Università degli Studi di Milano-Bicocca

Dipartimento di Biotecnologie e Bioscienze

Dottorato di ricerca in Biologia e Biotecnologie

XXXI Ciclo

GLUTAMATE, A NUTRIENT AT THE CROSS-ROAD OF CARBON AND NITROGEN ASSIMILATION

Marco Gnugnoli

Anno Accademico 2017/2018

SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Biotecnologie e Bioscienze

Dottorato di Ricerca in / PhD program Biologia e Biotecnologie Ciclo / Cycle XXXI

Curriculum in (se presente / if it is) Biologia Sintetica e dei Sistemi

GLUTAMATE, A NUTRIENT AT THE CROSS-ROAD OF CARBON AND NITROGEN ASSIMILATION

Cognome / Surname Gnugnoli Nome / Name Marco

Matricola / Registration number 736453

Tutore / Tutor: Prof. Marco Vanoni

Cotutore / Co-tutor:
(se presente / if there is one)

Supervisor:
(se presente / if there is one)

Coordinatore / Coordinator: Prof.ssa Paola Branduardi

ANNO ACCADEMICO / ACADEMIC YEAR 2017-2018

Summary

Abstract

Every living organism needs to tightly regulate its growth depending on the environment that surrounds it. This is true from the single-celled microorganisms to the mammals. Failure to correctly respond to environmental changes can be dangerous for the correct homeostasis of the cell. For this reason, the molecular basis and the regulation of many signalling pathways are conserved from bacteria to man and understanding this regulation is critical both for basic biology and to its application, from fermentations to medical science.

The work contained in this thesis is focused on the budding yeast *Saccharomyces cerevisiae* and on the analysis of its metabolism at the crossroad between carbon and nitrogen. *S. cerevisiae* is able to grow on different nitrogen sources, which can be classified as preferred or non-preferred. Not all the good nitrogen sources affect cellular processes in the same way. Glutamate is a very important amino acid and a good nitrogen source, being involved as nitrogen donor in several biosynthetic processes and possessing a carbon backbone (alpha-ketoglutarate), which can enter the TCA cycle.

In the first chapter, we investigated growth, metabolism and transcriptional profiles of yeast cells grown in minimal media supplemented with either ammonium or glutamate, both considered good nitrogen sources. During the exponential phase, cells using glutamate as a nitrogen source have a larger cell size compared to cells grown in presence of ammonium. In stationary phase, glutamate-supplemented accumulate higher biomass levels. We refer to this behaviour as “enhanced growth”. In glutamate-grown cells, the deepest transcriptional and metabolic rearrangement takes place after glucose exhaustion and show a profound alteration of the metabolism of storage molecules during growth on glutamate with an accumulation of trehalose and fatty acids, which correlate with a higher stress resistance. Flux balance analysis simulations with a core model of yeast metabolism correctly predicted the optimal growth yield and computational analysis of the flux distribution identified a different allocation of oxygen as responsible for the observed interplay between ethanol and glutamate.

Simulations using the model mentioned above suggest that growth on media containing glutamate as the only carbon and nitrogen source is possible. Despite these considerations, a wild-type *S. cerevisiae* strain is not able to grow in these conditions. To understand the reasons underlying this growth inability, in chapter 2 we applied a laboratory evolution approach and isolated 4 different mutants able to grow in the presence of just glutamate, vitamins and supplements. Sequencing of the mutant clones showed that they all carry mutations affecting the Ras/cAMP/PKA pathway. These results show that modification of the carbon sensing pathways can alter metabolism, allowing yeast to utilize glutamate in a way that the wild type strain

cannot. Transcriptomics analysis of the more interesting mutant revealed a general up-regulation of the biosynthesis of amino acids and nucleotides, as well as an enhanced expression of plasma membrane transporter genes, including glutamate permeases. Together with the enhancement of energy producing pathways, like fatty acids β-oxidation, we propose that these changes are among the driving forces in the adaptation of the evolved mutants.

In conclusion, our integrated analysis allowed us to demonstrate how glutamate cellular fate is strongly interlinked both with carbon and nitrogen metabolism and sensing and offers an example of combination of different techniques, which is able to deliver a system-level interpretation of biological data.

Introduction

Regulation of cellular growth

Every cell that exists, from the smallest single celled bacteria to the most specialized cell in our body, is not an island. Cells are, in fact, living in an environment and every cellular aspect needs to be tightly regulated in order to react to endogenous stimuli of different origin.

One of the fundamental cellular properties is the ability of the cell to consume and synthesize molecules to sustain life, grow in size, replicate its genome and organelles and divide. The regulation of this many aspects of growth is crucial for the maintenance of the homeostasis of the cell and of the multicellular organism the cell may be part of.

For this reason, different selective pressures led to the development of different regulatory networks depending on the type of organism and on its habitat.

Eukaryotic cells that form a multicellular organism respond primarily to hormones and growth factors, which can come either from the same cell (autocrine), from the surrounding cells (paracrine) or from distant organs (endocrine). For this reason, mammalian cells separated from their organism cannot be cultivated in vitro, even when all the nutrient they need are present, if growth-factor containing serum is not present. In fact, the ability to grow in the absence of these kind of external stimuli is a feature of cancer cells (Hanahan and Weinberg, 2011).

On the other hand, microorganisms respond primarily to the availability of nutrients to decide if and how to grow and divide.

The work in this thesis focuses on the regulation of growth by nutrients in the budding yeast *Saccharomyces cerevisiae*, a unicellular microorganism widely used both as a biotechnological tool (from the classical fermentations to the production of metabolites and heterologous proteins) and as a model organism for the study of different molecular mechanisms which are conserved in humans.

Since the regulation of cell growth between yeast and mammals depends on such different stimuli, it may not seem a good idea to use *S. cerevisiae* as a model organism. Anyway, the components of the signalling pathways are highly conserved, and various mutations in homologues of genes coding for yeast nutrient sensing pathways are associated with a deregulation of growth and tumorigenesis in humans (Bos, 1989) (Eltschinger and Loewith, 2016) (Faubert *et al.*, 2015).

S. cerevisiae metabolism

Yeast carbon metabolism is centred on glucose

Yeast is able to metabolize a lot of different carbon sources, mostly carbohydrates deriving from plants. This has been exploited for a long time by humans for production of food and beverage, at first in a spontaneous way and later with more consciousness of the fact that fermentation was the result of the action of a living organism.

Glycolysis: from glucose to pyruvate

To perform the reaction that we use for all those processes yeast must break down sugars to obtain energy and building blocks for anabolic reactions. The first step toward this objective is a process known as glycolysis, which happens without the requirement for the presence of oxygen and produces energy storage molecules like the nucleotide triphosphate ATP and reduces molecules of the coenzyme NAD to NADH.

The starting molecule of glycolysis is glucose, a monosaccharide that is yeast favourite carbon source and is, therefore, able to regulate a variety of cellular processes by the action of sensing and signal transduction systems (see Effect of the signal transduction pathways). Glucose can be directly imported in the cell via specific transporters (see Glucose) or derived from more complex or different sugars. The first step of glycolysis is allowed by enzymes called hexokinases, which attach a phosphate moiety from ATP to a glucose molecule to yield glucose 6-phosphate (Figure 1). Yeast possesses three hexokinases, encoded by *HXK1*, *HXK2* and *GLK1* genes. Hxk2 gene is the main hexokinase in cells growing on glucose (Bianconi, 2003).

Glucose 6-phosphate is later converted, by the action of Pgi1, to fructose 6-phosphate, which is, in turn, phosphorylated to fructose-1,6-bisphosphate. This reaction consumes one more molecule of ATP and is catalysed by the gene products of *PFK1* and *PFK2*, which are the two subunits of phosphofructokinase. The aldolase encoded by *FBA1* and the triose phosphate isomerase Tpi1 allow the production of 2 molecules of glyceraldehyde-3-phosphate (Compagno, Dashko and Piškur, 2014).

This ends the ATP-consuming part of glycolysis, which is followed by the second part, where energy storage molecules are produced. The reactions catalysed by Tdh1/2/3, Pgk1, Gpm1, Eno1/2 and Pyk2/Cdc19 lead to the production of 2 molecules of pyruvate for every molecule of glucose. During these reactions, 4 molecules of ATP and 2 of NADH are generated (Compagno, Dashko and Piškur, 2014).

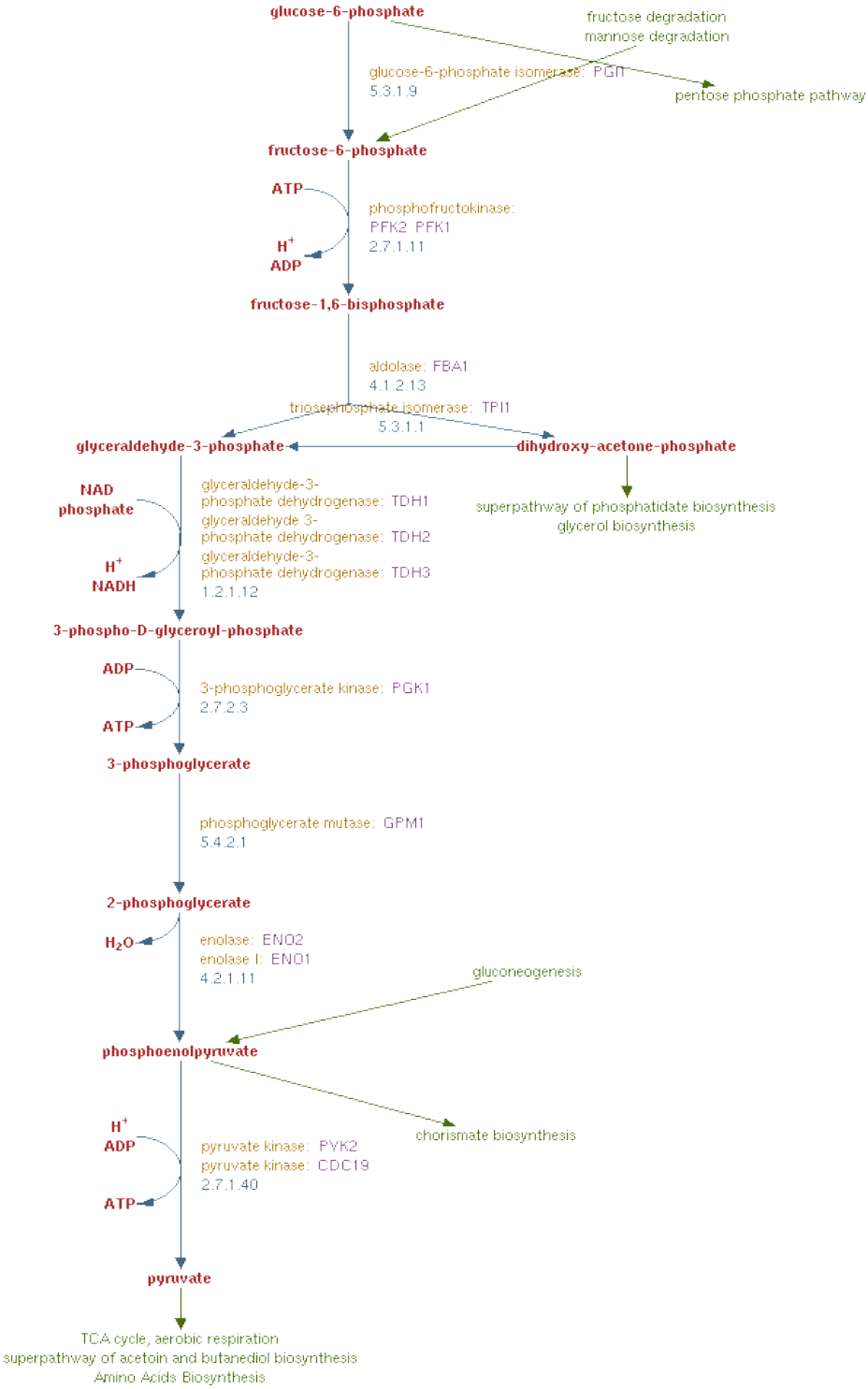


Figure 1 – Scheme of glycolysis (from yeastgenome.org)

The many fates of pyruvate reflect different growth strategies

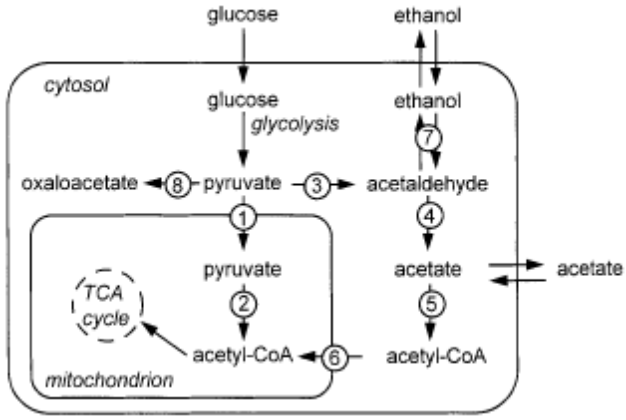


Figure 2 – Main reactions associated with pyruvate metabolism (Pronk, Yde Steensma and Van Dijken, 1996)

The pyruvate generated this way is the starting point of many biochemical pathways (summarized in Figure 2), that eventually lead to the re-oxidation of NADH molecules to NAD^+ (Pronk, Yde Steensma and Van Dijken, 1996).

One particular pathway that start from pyruvate has been largely exploited by mankind over the centuries: alcoholic fermentation. In this process, one molecule of pyruvate is decarboxylated to acetaldehyde, which is, in turn, converted into ethanol in a NADH-consuming reaction catalysed by one of the alcohol dehydrogenases isozymes encoded by *ADH1, ADH3, ADH4* and *ADH5* (de Smidt, du Preez and Albertyn, 2008).

The other possible metabolic fate of pyruvate is respiration. This process takes place in the mitochondria, so it requires the transport of pyruvate in the organelle and its conversion to acetyl-CoA, catalysed by the pyruvate dehydrogenase complex, composed by Lat1-Pda1-Pdb1-Lpd1-Pdx1 (Pronk, Yde Steensma and Van Dijken, 1996). There, the acetyl-CoA generated goes through 10 reactions (the tricarboxylic acid, or TCA cycle), shown with the corresponding enzyme-coding genes in Figure 3.

At the end of these reaction acetyl-CoA is transformed into oxaloacetic acid, which ends up reacting with a new acetyl-CoA molecule in the first reaction, catalysed by citrate synthase, therefore closing the cycle. As a final result of the TCA cycle, 2 molecules of NAD^+ are reduced to NADH that add up to the molecule formed during glycolysis (Compagno, Dashko and Piškur, 2014). The electrons carried by those NADH molecules are then transferred to the electron transport chain, in a process that allows for the production of ATP and the re-oxidation of all the NADH molecules. Moreover, in the reaction catalysed by succinate dehydrogenase (also known as respiratory chain

complex II), electrons are transferred to a ubiquinone molecule, reducing it to ubiquinol (Rutter, Winge and Schiffman, 2010).

Another important function of the TCA cycle is the generation of precursor molecules for various biosynthetic processes. In particular, oxaloacetic acid is required for the biosynthesis of lysine and aspartate (Jones and Fink, 1982) and 2-oxoglutarate (also known as α-ketoglutarate) is required for the biosynthesis of glutamate (Cooper, 1982). Because the reactions forming and degrading glutamate are central in yeast nitrogen metabolism (see p.16), this molecule is the interface between carbon and nitrogen metabolism.

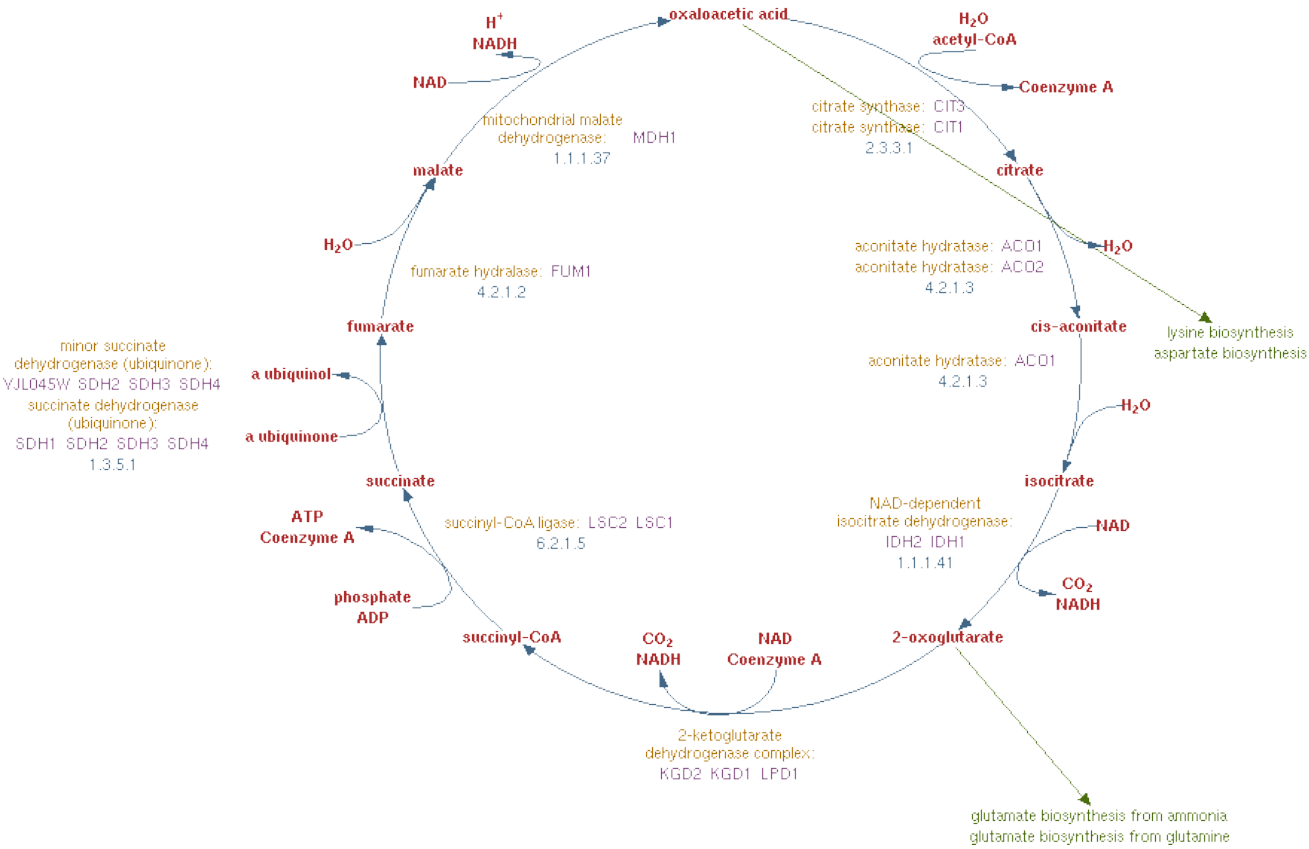


Figure 3 - TCA cycle (from yeastgenome.org)

The next step is the transfer of the electrons stored in the NADH molecules to a series of proteins and protein complexes located in the mitochondrial matrix, called the electron transport chain (Figure 4). The first component of the electron transport chain

is the NADH:ubiquinone oxidoreductase encoded by *NDE1*, which carries out the function of mammalian complex I, reducing a ubiquinone molecule to ubiquinone. In contrast with complex I, Nde1 does not pump protons across the membrane. As mentioned in the last paragraph, complex II also carries out the same reactions with the electrons

taken from succinate in the TCA cycle. Complex III catalyse the opposite reaction, transferring the electrons from ubiquinol to cytochrome c. This first cycle of reduction and re-oxidation of ubiquinone, called Q cycle, leads to the translocation of protons to the mitochondrial intermembrane space. A similar cycle is the result of the action of complex IV, which oxidizes cytochrome c transferring the electrons to the final acceptor, O_2. The protons translocated with these two reactions then pass through the F1F0 ATP synthase which is able to generate ATP. In yeast, the total amount of ATP generated through glycolysis and respiration of a glucose molecule is 16 ATP (Bakker *et al.*, 2001).

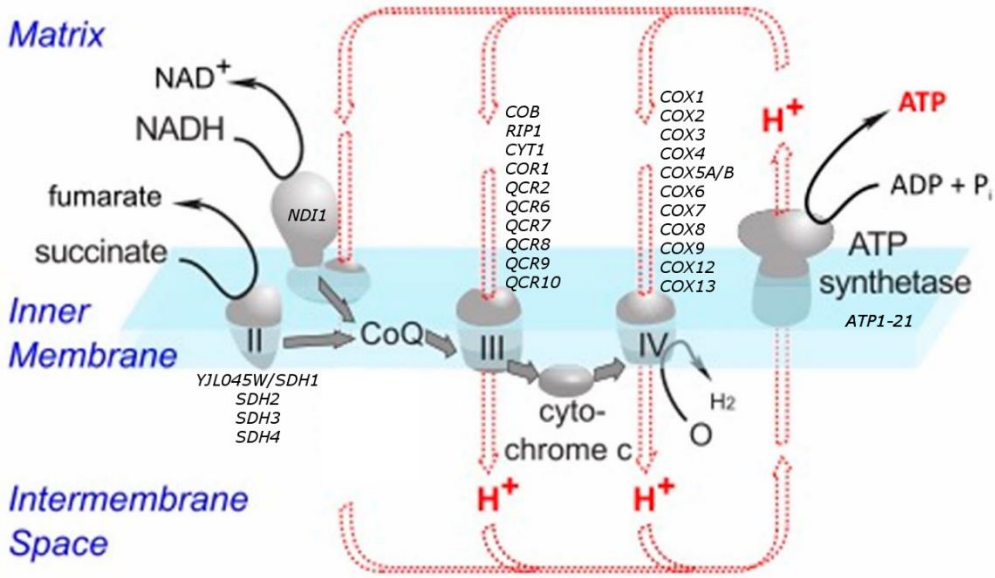


Figure 4 - Electron transport chain, general scheme and protein-coding genes
(adapted from Van Der Bliek, Sedensky and Morgan, 2017)

Analysing the number of ATP molecules produced by alcoholic fermentation and respiration, it is evident that most efficient pathway is the latter. Despite this, yeast is mostly known for its ability as a fermenter, even in the presence of oxygen.

The reason behind this weird behaviour is the so-called “Crabtree effect”. Named after its discoverer, it is the fact that *S. cerevisiae* (and the other Crabtree-positive yeasts), even during aerobic growth, prefer fermentation over respiration, as long as a large quantity of glucose is present in the growth medium.

This has been explained as a precise growth strategy which has conferred several advantages to *S. cerevisiae*, called “make-accumulate-consume” (Piškur *et al.*, 2006). This strategy is mostly due to the repression exerted by glucose and other C_6 sugars on respiration genes (see p. 31). It allows yeast to quickly consume a large quantity of glucose, therefore subtracting from its competitors. Moreover, the ethanol produced inhibits the growth of many other microorganisms, while *S. cerevisiae* is resistant. Finally, if oxygen is present, yeast can utilize the ethanol previously produced, giving rise its characteristic “diauxic growth” (Figure 5).

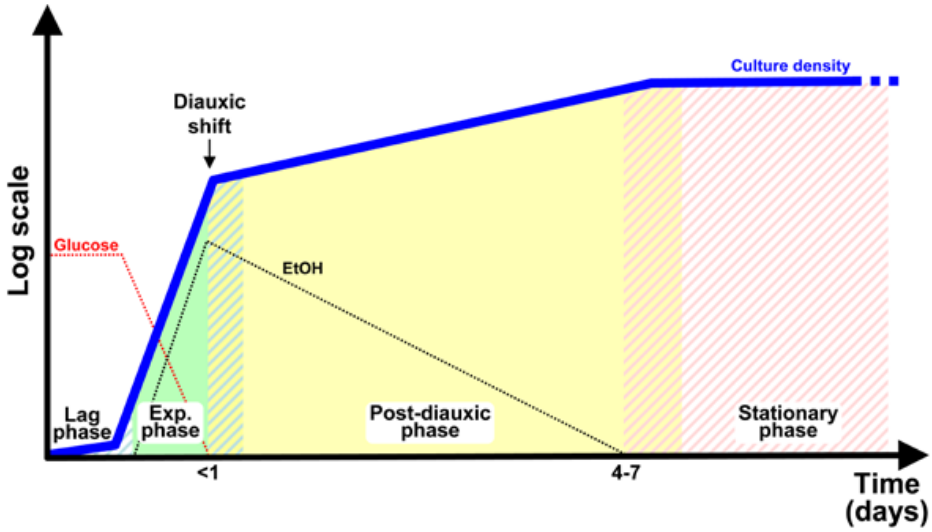


Figure 5 - Typical growth curve of a yeast culture growing on a glucose-containing medium. S. cerevisiae cells, after an initial lag phase, grow rapidly on glucose and produce ethanol, which is later consumed after the diauxic shift (Busti *et al.*, 2010).

Yeast Ethanol consumption is due to the catalysis operated by Adh2 enzyme, which converts the alcohol to acetaldehyde, that is, in turn, transformed into acetate. Both the enzymes are dehydrogenases and reduce NAD^+ to NADH. Acetyl-CoA is then synthesized from acetate by Acs1 and Acs2 and enters the TCA cycle, fuelling the electron transfer chain.

Interestingly, human tumour cells show a similar metabolism of glucose, with an accelerated glycolytic flux and the largest portion of ATP formed by substrate-level phosphorylation and not by oxidative phosphorylation. Fermentation by cancer cells (also known as Warburg effect) has lactate as a final result instead of ethanol. So, despite many differences in the regulation of the biochemical pathways, yeast is a very interesting mode of human deregulated cancer metabolism (Legiša, 2014).

Anabolism and anaplerotic reactions

When yeast cells are growing on non-fermentable carbon sources, like ethanol or other C_2 or C_3 compounds, they need to synthesize hexoses for biosynthetic purposes. This aim is reached via the gluconeogenetic pathway. Gluconeogenesis is, essentially, the reverse of glycolysis and share with it many enzymes. The few different enzymes are those catalysing the irreversible reactions, that involve ATP production or consumption (Compagno, Dashko and Piškur, 2014). The conversion of pyruvate into phosphoenolpyruvate (with oxaloacetate as intermediate) is mediated by the subsequent action of pyruvate carboxylase, encoded by *PYC1* or *PYC2,* and phosphoenolpyruvate carboxykinase, encoded by *PCK1* (Pronk, Yde Steensma and Van Dijken, 1996).

Yeasts also need to produce molecules to replace missing intermediates in the TCA cycle that are removed for biosynthetic purposes. This necessity is satisfied by the so-called anaplerotic reactions. The reaction catalysed by Pyc1/2 is important in fuelling the cycle with oxaloacetate (Pronk, Yde Steensma and Van Dijken, 1996). Another relevant pathway for this purpose is the glyoxylate cycle. This pathway is composed, for the first three, of shared reactions with the TCA cycle, from oxaloacetate to isocitrate. The last reaction takes place in the cytoplasm, where citrate is previously exported (Strijbis and Distel, 2010). There, instead of converting isocitrate to α-ketoglutarate, the isocitrate lyase encoded by *ICL1* converts it to succinate and glyoxylate (Fernández, Moreno and Rodicio, 1992), which is in turn imported into the peroxisome, where the malate synthase Mls1 condensate Acetyl-CoA and glyoxylate into malate (Hartig *et al.*, 1992). The pentose phosphate pathway is another important metabolic route which subtract glucose-6-phosphate from glycolysis for biosynthetic purposes. In particular, the hexose phosphate goes through a series of reactions shown in Figure 6, that lead to the production of intermediates required for the anabolism of amino acids and nitrogen bases and NADPH, needed both for providing the energy

required for metabolic reactions and for oxidative stress resistance (Compagno, Dashko and Piškur, 2014).

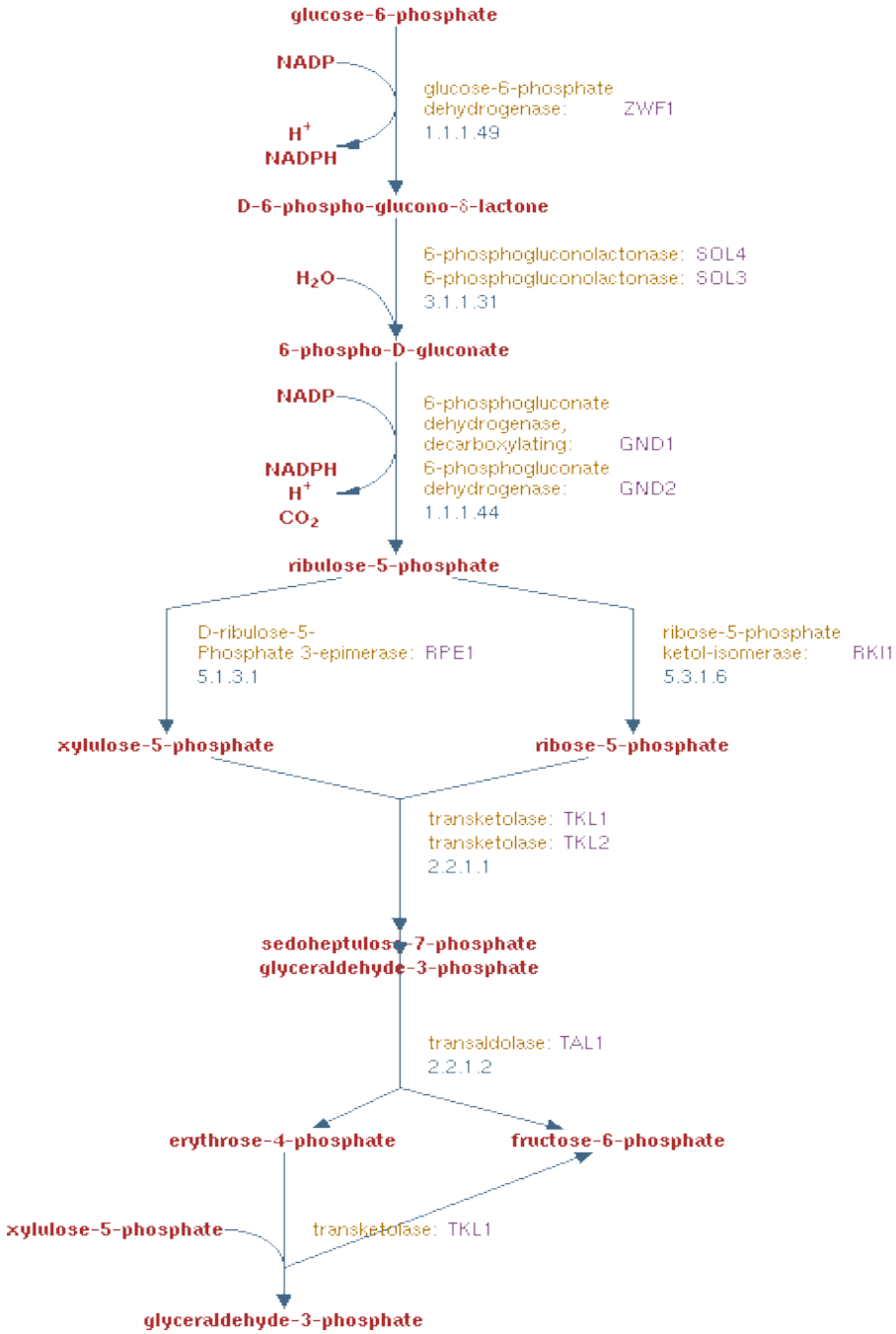


Figure 6 - Pentose phosphate pathway (from yeastgenome.org)

Nitrogen metabolism revolves around glutamate and ammonia

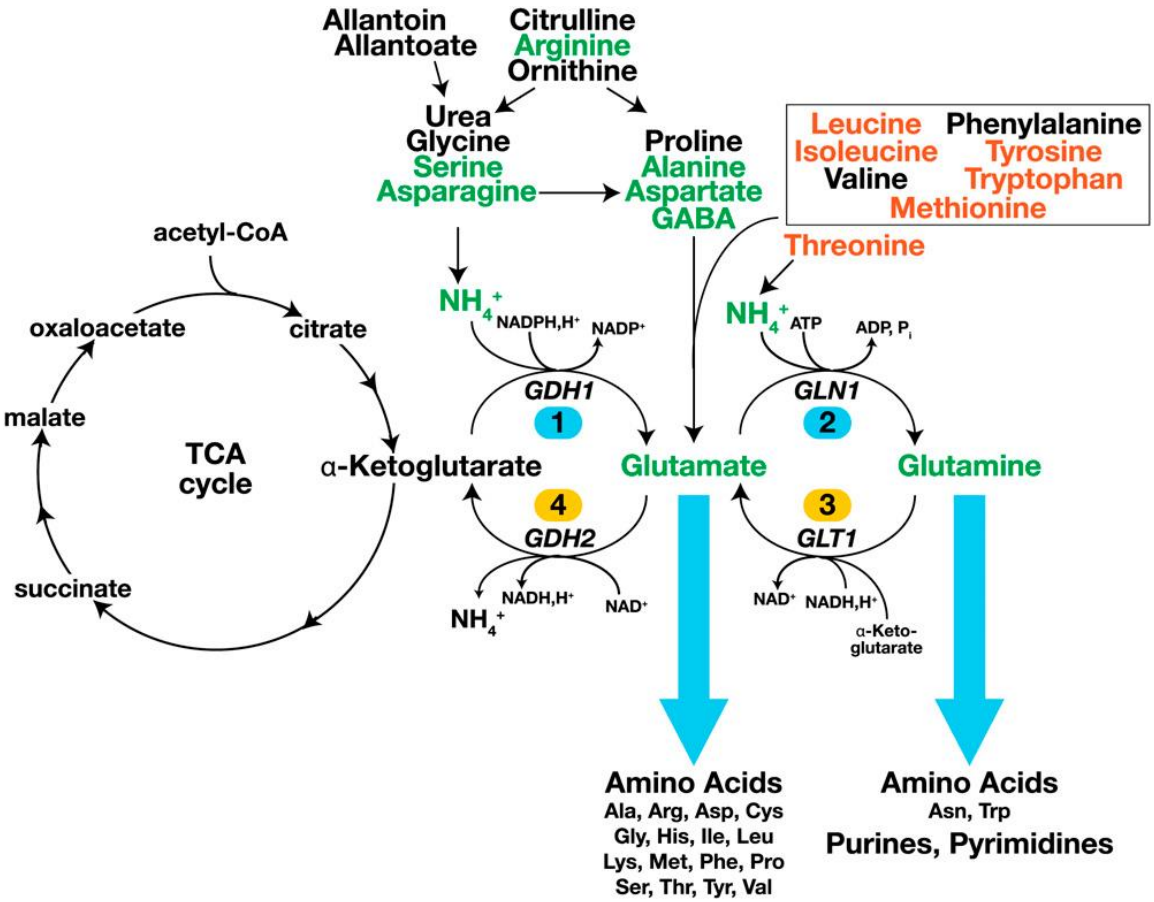


Figure 7 - Assimilation and utilization of the main nitrogen sources in S. cerevisiae (Ljungdahl and Daignan-Fornier, 2012)

Yeast natural habitat is composed by fruits and vegetables that undercome decomposition. This environment is particularly enriched in compounds that serve as nitrogen sources. For this reason, the metabolic routes that yeast uses to incorporate that nitrogen have to account for the enormous heterogeneity of the possible nitrogen containing compounds (Figure 7). The type of nitrogen sources that yeast can metabolize are most of the amino acids, urea, compounds derived from purines and the nitrogen base uracil. Each of this substances is able to sustain yeast growth at a different growth rate, and some can only allow yeast to perform a limited number of cell duplications, therefore it is not completely true to affirm that they are able to be used as nitrogen sources by yeast (Cooper, 1982).

The study of the different cellular degradation systems for nitrogen containing compounds has been performed mostly by isolation of mutants lacking specific

enzymes in the various catabolic pathways or genes coding for protein involved in the control of nitrogen metabolism.

From a lot of possible nitrogen sources derive a lot of catabolic pathways. This leads to a difficult classification of this pathways. Luckily a lot of this metabolic branches converge on common substances and this allow to group them according to the final product of their degradation.

The first group of pathways is the one that generate ammonia as the final product. This includes the allantoin, urea and asparagine degradation pathways. The ammonium generated in this way must be converted to glutamate by NADP-dependent glutamate dehydrogenases (Gdh1 and Gdh3). The other pathways converge directly on glutamate. For this reason, glutamate and ammonia are central in yeast nitrogen metabolism (Cooper, 1982), and this is one of the reasons that led us to investigate the differences observed between these two compounds used in culture media. The relationship between them is what I will mostly focus on in the next paragraph.

Glutamate and ammonia are the interface between nitrogen compounds anabolism and catabolism. They can sustain yeast growth as sole nitrogen sources and are the most important nitrogen donors in the cellular biosynthetic reactions, together with glutamine. The reactions that convert ammonia and α-ketoglutarate into glutamate and vice versa is catalysed by glutamate dehydrogenases (GDHs). Yeast possesses 3 different GDHs, characterized using a different cofactor.

GDH2 catalyses the conversion of glutamate to ammonia, according to the reaction:

$$\text{L-glutamate} + NAD^+ \leftrightarrows \alpha\text{-ketoglutarate} + NADH + NH_4^+$$

GDH1 and GDH3 are the two NADP-dependent isoforms responsible for the opposite reaction:

$$\alpha\text{-ketoglutarate} + NADPH + NH_4^+ \leftrightarrows \text{L-glutamate} + NADP^+$$

The other option yeast cells have for the interconversion of ammonia and glutamate is the reaction catalysed by the synergic action of glutamine synthase GLN1:

$$\text{L-glutamate} + ATP + NH_4^+ \rightarrow \text{L-glutamine} + ADP + P_i$$

and glutamate synthase GLT1:

$$\text{L-glutamine} + \alpha\text{-ketoglutarate} + NAD(P)H \rightarrow 2\ \text{L-glutamate} + NAD(P)^+$$

which, in the end, converts α-ketoglutarate and ammonia to glutamate (Cooper, 1982).

None of these pathways is essential, as shown by the fact that the single *gdh1Δ, gdh3Δ* and *glt1Δ* mutants, as well as the double mutants, are able to grow on a synthetic medium containing ammonia as a nitrogen source, even if with different growth defects depending on the mutation and on the carbon source (DeLuna *et al.*, 2001) (Avendano *et al.*, 1997). The triple deletion mutant is not able to grow in such condition, while it is able to grow when glutamate is provided as nitrogen source (Campero-Basaldua *et al.*, 2017).

This indicates a partial redundancy of pathways for ammonia assimilation. As said before, the deletions of the enzymes have different effects during growth on different carbon sources. This difference is particularly evident in the case of the two paralogs Gdh1 and Gdh3. *GDH3* gene expression is induced in the presence of ethanol and repressed during growth on glucose, while GDH1 levels do not change. Another difference between the two isoforms is that Gdh1 consumes α-ketoglutarate at a higher rate (DeLuna *et al.*, 2001). We can conclude that the fact that Gdh1 is more important for growth on fermentable carbon sources, while Gdh3 with non-fermentable carbon sources is probably due to the different availability of α-ketoglutarate and energetic reasons.

Biosynthesis of amino acids

Every cell needs amino acids to build proteins for sustaining life. Unlike mammalian cells, yeast is able to synthesize all the amino acids required for protein synthesis starting from the backbone derived from a carbon source and ammonia, with the reactions described in the last paragraph.

Amino acids can be divided into different families that are derived from a common precursor amino acid, summarized in Table 1.

Family	AAs
glutamate	glutamate, glutamine, arginine, proline, lysine
aromatic	phenylalanine, tyrosine, tryptophan
serine	serine, glycine, cysteine, and methionine
aspartate	aspartate, asparagine, threonine, cysteine, methionine
pyruvate	alanine, valine, leucine, isoleucine

Table 1 - Biosynthetic family of amino acids

Biosynthesis of nucleotides and energy metabolism

Another important class of nitrogen-containing compounds is nucleotides. They serve different roles in the cell: they form nucleic acids, serve as energy supply for the different cellular processes (ATP), they are used as signalling molecules (GTP, cAMP) and they are important components of cofactors (NAD, FAD, Coenzyme A) (Ljungdahl and Daignan-Fornier, 2012).

Nucleotides follow different biosynthetic pathways for purines and pyrimidines synthesis which include both salvage pathway for the reutilization of components of the growth medium or of the cellular catabolism and *de novo* synthesis from amino acids and carbohydrates (Figure 8).

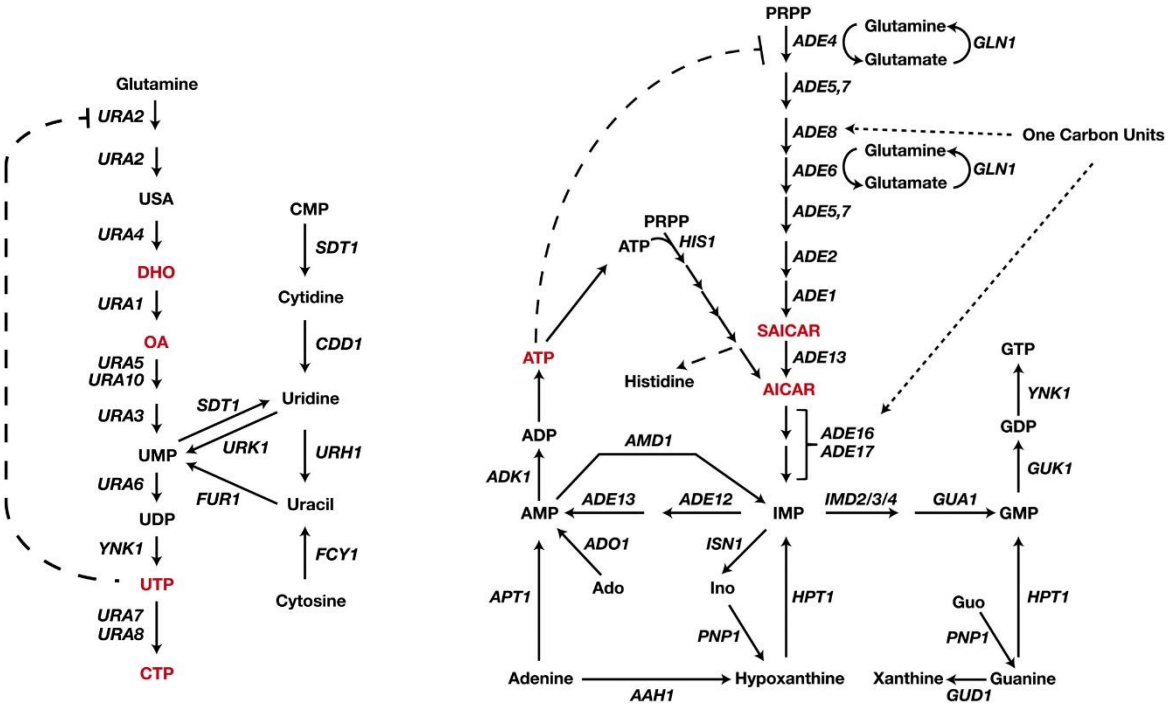


Figure 8 - De novo and salvage synthesis of pyrimidines (left) and purines and histidine (right) (Ljungdahl and Daignan-Fornier, 2012)

As shown in the figure, nitrogen metabolism and in particular glutamate and glutamine, have an important role as precursor or donor for the synthesis of both types of nucleotides. In addition, phosphate is also a crucial nutrient for the synthesis of different nucleotides, as shown by the fact that among mutants carrying overexpression of the *PHO5* gene, encoding for yeast acid phosphatase, the loss of function *adk1* and *ado1* mutants were found (Gauthier *et al.*, 2008). These genes are, respectively, the adenylate and adenosine kinase, responsible for the synthesis of

AMP and ADP. These enzymes, together with those involved in metabolic reactions that produce ATP are really important because of the importance of adenosine nucleotides (collectively called AXP) in the energy metabolism and sensing of the cells.

This reflects in the fact that a commonly used parameter to measure the energetic state of the cell is the so called adenylate energy charge, calculated with the following equation:

$$EC = ([ATP] + 0.5\,[ADP])/([ATP] + [ADP] + [AMP])$$

When all adenine nucleotide pool is in form of AMP the energy charge (EC) is zero, and the system is completely discharged (zero concentrations of ATP and ADP). With only ADP, the energy charge is 0.5. If all adenine nucleotide pool is in form of ATP, the EC is 1. Therefore, the EC value could be a measure of the cell's ability to carry out the anabolic and maintenance reactions required to grow and to remain alive (Atkinson and Walton, 1967) and the regulation of these reactions involving phosphate is crucial for the maintenance of energy homeostasis.

Glycogen and trehalose are yeast reserve carbohydrates

Just like us, yeast cells must store the extra energy obtained by nutrient (especially glucose). Amongst the various ways of storage, two fundamental molecules are glycogen and trehalose. Aside from being similar in some ways, they are synthesized in different moments of yeast growth and their function is not the same.

The starting molecule for production of both trehalose and glycogen is UDP-glucose, which is then polymerized in different ways by the action of different enzymes (Figure 9).

Glycogen is a branched polysaccharide with a high molecular mass. It is composed of α-(1, 4)-glucosyl chains with α-(1, 6)-linkages. The average length of each linear chain is 13 glucose molecules (François, Walther and Parrou, 2012). Biosynthesis of glycogen is similar to that of many biomolecules of high molecular mass. It is composed of 3 fundamental steps: initiation, elongation and ramification (Francois and Parrou, 2001).

The initiation step is performed by a protein called glycogenin, which possess a self-glycosylating activity and reacts with UDP-glucose, forming a short α-(1,4)-glucosyl chain, by formation of a covalent bond with a tyrosine residue. Yeast glycogenin is encoded by *GLG1* and *GLG2* genes (François, Walther and Parrou, 2012). While the presence of these genes is not essential for glycogen biosynthesis, the double *glg1 glg2* mutant shows a random accumulation of glycogen only in some colonies. Therefore, the existence of alternative initiation routes for glycogen biosynthesis is probably true, but glycogenin presence ensures a good level of glycogen production in every cell (Torija *et al.*, 2005).

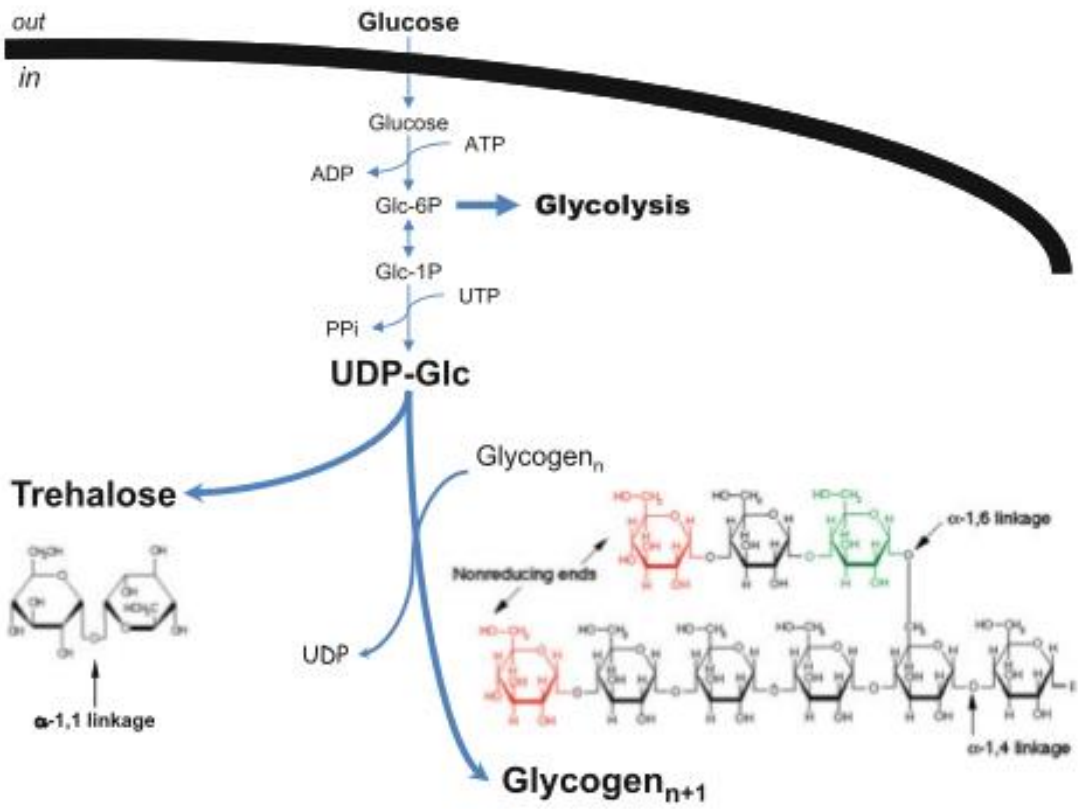


Figure 9- Structure and metabolic routes for trehalose and glycogen biosynthesis (from François, Walther & Parrou, 2012)

The elongation step is operated by the glycogen synthase, which is present in yeast as the two isoforms Gsy1 and Gsy2, which are very similar proteins and catalyse the polymerization of glucose units with α-(1,4)- glycosidic bonds. The importance of the two isoforms is different in cells growing in different carbon sources, with Gsy2 being responsible for most of the glycogen production during growth on glucose, and the two being equally important in cells growing on galactose (François, Walther and Parrou, 2012).

The branching step is due to the action of the amylo (1,4)→(1,6)-transglucosidase encoded by *GLC3*, which catalyse transfer of 6-8 glucose residues from the end of a linear chain to form a α-(1,4)- glucosidic bond with an internal residue (Francois and Parrou, 2001). Branching activity is fundamental for glycogen biosynthesis, as shown by the deletion of *GLC3,* which strongly reduce the polysaccharide accumulation (Thon *et al.*, 1992).

Glycogen breakdown occurs via two distinct mechanism for the cytosolic and vacuolar compartments. Cytosolic glycogen is degraded by the action of glycogen phosphorylase Gph1 and the debranching enzyme Gdb1. The former catalyses the removal of glucose-1-P from the linear α-(1,4) chain, while the latter is able to transfer maltosyl or maltotriosyl molecules from a branch to a linear chain and to remove the α-(1,6) bond. Gph1 is activated by phosphorylation, the players involved still not being completely elucidated (see PKA). Vacuolar glycogen is imported in the compartment during growth on glucose, and constitute a reservoir protected from cytosolic degradation. The breakdown of this pool is due to the action of the gene product of *SGA1*, which encodes an amylo-(1,4),(1,6) glucosidase, whose reaction yields glucose as final product. Expression of *SGA1* is induced during stationary phase or starvation and during sporulation (François, Walther and Parrou, 2012).

Trehalose, on the other hand, is a much simpler molecule, being a disaccharide composed by two glucose molecules connected by a α-(1,1) bond (Figure *9*). The conversion of UDP-glucose to trehalose is performed by a protein complex containing the two trehalose synthases, encoded by *TPS1* and *TPS2*, together with the structural proteins encoded by *TPS3* and *TLS1* (Figure 10). The reaction proceeds in two steps, the first being the conversion of UDP-glucose to trehalose-6-P (catalysed by Tps1) and the subsequent removal of the phosphate moiety by the trehalose-6-P phosphatase Tps2.

Trehalose degradation is due to the action of hydrolytic enzymes called trehalases, that are able to catalyse a reaction yielding 2 molecules of glucose from every molecule of trehalose. Yeast possesses two types of trehalase activities: neutral (with a pH optimum around 7) and acid (optimal pH=4) (François, Walther and Parrou, 2012). Cytosolic trehalose catabolism is carried out by the neutral trehalase activity, which is mostly due to the action of the enzyme encoded by the *NTH1* gene. In fact, deletion of this gene abolishes intracellular trehalose metabolization in most growth conditions. Nth1 protein, like many others involved in storage compound metabolism, is regulated by PKA-dependent phosphorylation (see PKA). *S. cerevisiae* possesses a second neutral trehalase, encoded by *NTH2* and paralog of *NTH1*, whose enzymatic activity was doubtful for a long time. In more recent times, it was shown that its gene product is actually responsible for a "residual neutral trehalase activity" in a *nth1Δ* strain during particular growth conditions, such as stationary phase (Jules *et al.*, 2008) and recovery from stress (Garre and Matallana, 2009).

The acid trehalase activity of yeast is due to the action of the protein Ath1, which is localized to both the vacuoles and at the periplasm and cell wall. While it has been established that Ath1 is active at the surface of the cell and that its presence is required for extracellular trehalose utilization (Jules *et al.*, 2004), the vacuolar function of the protein is still to be elucidated. Some have hypothesized that the vacuolar acid

trehalase is just sequestered there to avoid degradation of trehalose protecting the cells surface (Eleutherio *et al.*, 2014).

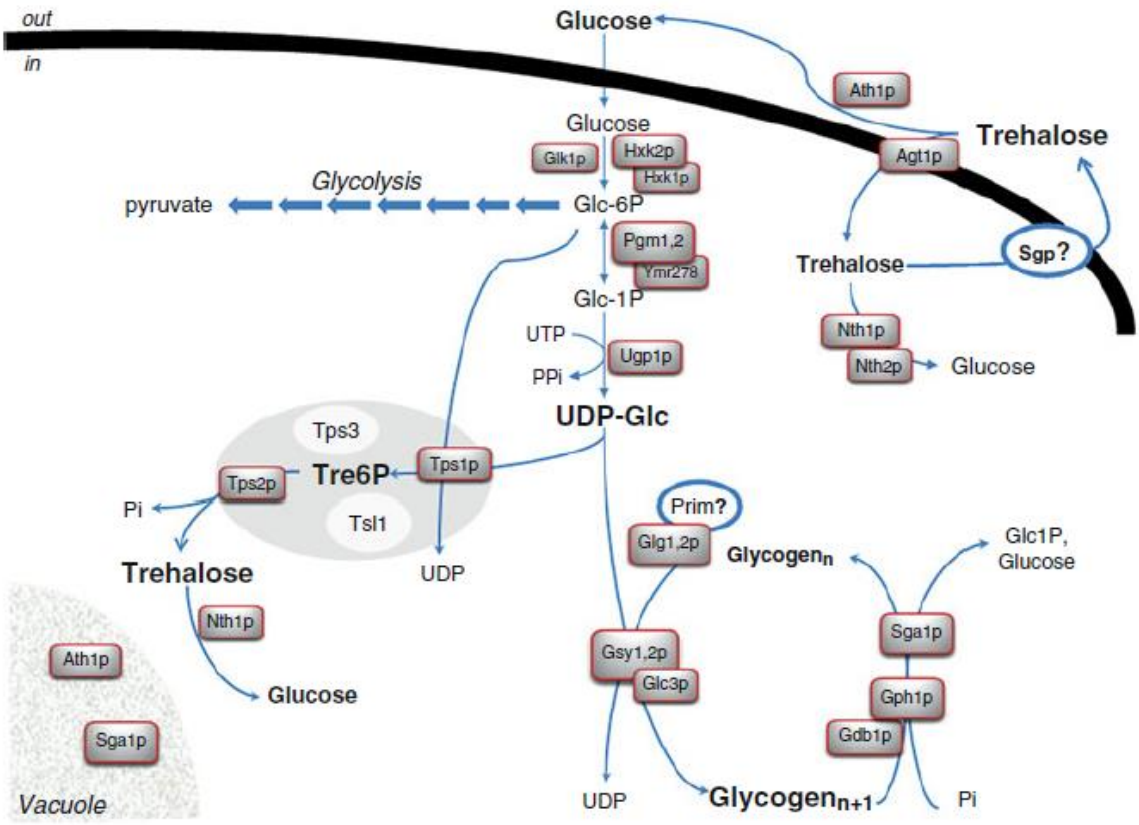


Figure 10 - Biosynthesis and degradation of trehalose and glycogen (François, Walther and Parrou, 2012)*)*

Storage lipids and lipid droplets

Yeast has been extensively used as a model for the study of lipid metabolism. Like other living organisms, yeast metabolism produces and utilize fatty acids (FAs), glycerolipids, glycerophospholipids, sterols and their derivatives, sphingolipids, prenol lipids, glycolipids, and polyketides.

Lipids serve various functions in the cell, two of the most important being forming the biological membranes (sphingolipids and phospholipids) and serving as energy storage molecules (triglycerides and steryl esters). Both these kind of molecules derive from FAs, which can derive from biosynthesis, breakdown of more complex lipids and uptake from the environment (Klug and Daum, 2014).

Fatty acids can be either free (FFAs) or assembled in more complex lipids. They are carboxylic acids which possess a carbon chain with a defined length and can be fully saturated or have a certain degree of unsaturation (i.e. the number of double bonds in the carbon chain). Yeast have a defined composition of fatty acids, which was defined experimentally and is susceptible of changes depending on the growth condition of the cells. Generally speaking, yeast most abundant FAs are palmitoleic acid (C16, 1 unsaturation), palmitic acid (C16, fully saturated) and stearic acid (C18, 1 unsaturation) and the less abundant are myristic acid (C14, 1 unsaturation) and cerotic acid (C26, fully saturated) (Klug and Daum, 2014).

De novo synthesis of FAs is the most related to the topics of this thesis, because it is the route by which energy and elements extracted from other environmental nutrients are incorporated into lipids. It takes place in the cytoplasmic and mitochondrial compartments, and later in the endoplasmic reticulum (ER) for elongation and desaturation.

The first step in FAs synthesis is the reaction catalysed by acetyl-CoA-carboxylase:

$$ATP + \text{acetyl-CoA} + HCO3^{-} \rightleftharpoons ADP + \text{phosphate} + \text{malonyl-CoA}$$

Yeast possesses two isoforms of this enzyme: one cytoplasmic that uses Acetyl-CoA, encoded by *ACC1*, and another mitochondrial, encoded by *HFA1*.

The next steps are catalysed by two different fatty acid synthase (FAS) complexes in the cytoplasm and in the mitochondria. In particular, cytosolic (or type I) FAS complex is composed of two subunits formed, in turn by hexamers of Fas1 (β subunit) and Fas2 (α subunit) proteins. These proteins bear multiple enzymatic activities. In particular, Fas1 possess acetyl transferase, enoyl reductase, dehydratase and malonyl-palmitoyl transferase activities. Fas2 bears acyl carrier protein, 3-ketoreductase, 3-ketosynthase and phoshopantheteine transferase activities (Tehlivets, Scheuringer and Kohlwein, 2007). Mitochondrial (or type II) FAS complex, instead, is formed by six proteins that

only have one enzymatic activity each. In detail, Acp1 is the acyl-carrier protein, Cem1 is the β-ketoacyl-ACP synthase, *OAR1* encodes a 3-oxoacyl-acyl-carrier protein reductase, Htd2 functions as 3-hydroxyacyl- thioester dehydratase, Etr1 is an enoyl-ACP reductase and *PPT2* encodes a phosphopantetheine protein transferase (Klug and Daum, 2014). The final results of the action of both FAS complexes is the transfer of two carbons from a molecule of malonyl-CoA to a growing saturated acyl chain, as shown in Figure 11.

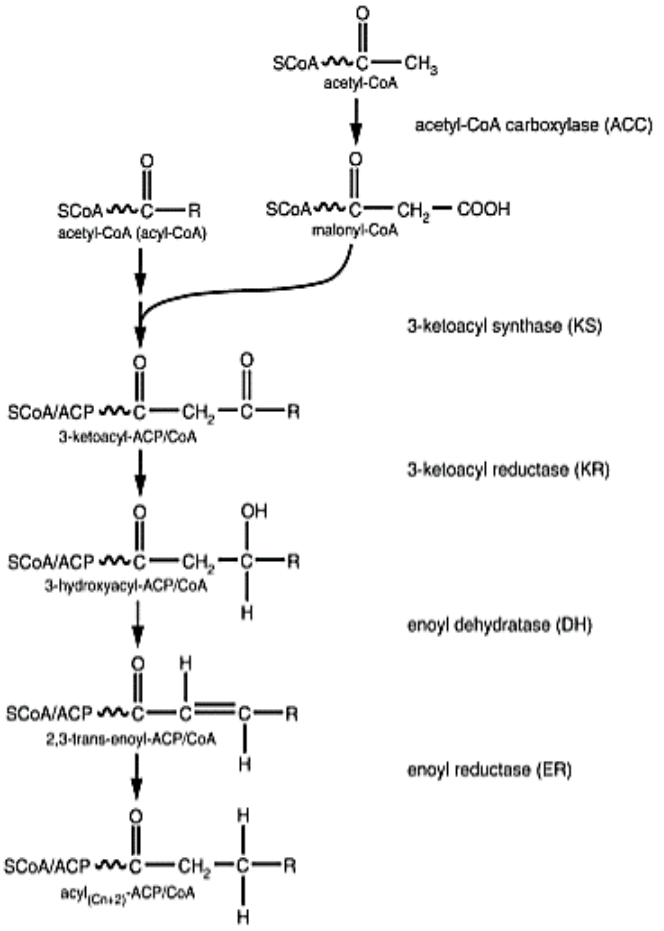


Figure 11 - Main reactions of fatty acid synthesis, from (Tehlivets, Scheuringer and Kohlwein, 2007)

Addition of more carbon atoms to the long chain fatty acid generated via these reactions takes place in the ER and requires the action of specialized proteins called elongases, encoded by *ELO1, ELO2* and *ELO3* genes (Rössler *et al.*, 2003). Finally,

introduction of double bonds is due to the desaturation reaction catalysed by the protein Ole1 (JE, VM and CE, 1989).

FFAs that are not transformed in the amphiphilic lipids that compose the biological membranes (sphingolipids and phospholipids) can be toxic for the cell. For this reason, they are transformed into bigger inert molecules: triacyclglycerols (TAGs), composed of three fatty acid chains linked to a glycerol molecule. The same goes for sterols, another class of membrane lipids synthetized from Acetyl-CoA and formed by four rings of carbon atoms with a side acyl chain and an hydroxylic group which serves as a hydrophilic moiety. These compounds can also cause lipotoxicity and are therefore converted into sterol esters (SEs) by esterification with a fatty acid (Klug and Daum, 2014).

Both these two types of molecules constitute yeast way of storing fats, that at the right moment can be used to obtain energy and to synthesize membrane lipids. They are stored in specialized cellular organelles, called lipid droplets (LDs). At first these particles were considered just reservoirs of the above mentioned neutral lipids, but they turned out to be very dynamic organelles, whose defects in multicellular eukaryotes give rise to a series of diseases (Kohlwein, Veenhuis and van der Klei, 2013). LDs are composed of an external monolayer of phospholipids which contain the TAGs and SEs and contain proteins as well, mostly enzymes of lipid metabolism.

For example, degradation of storage lipids is performed on the surface of lipid droplets, where TAGs are hydrolysed by lipases (encoded by *TGL3*, *TGL4* and *TGL5*) and SEs are attacked by the action of hydrolases Tgl1 and Yeh1 (Radulovic *et al.*, 2013).

Another important feature of LDs is their ability to interact with other organelles. The main example is their interaction with peroxisomes, which is associated with the β-oxidation of FAs. That is the process by which fatty acids are degraded to yield acetyl-CoA and a reduced acyl-CoA with 2 carbon atoms. This process requires the translocation of FAs into the peroxisome operated by the complex formed by the proteins Pxa1 and Pxa2 (for long chain fatty acids) or by Ant1 and Pex11 (for medium chain fatty acids). For the first group the activation with coenzyme A happens in the cytoplasm before the import, while for the latter it is performed inside the peroxisomal compartment by the medium chain fatty acyl-CoA synthetase Faa2. Subsequently, the enzyme encoded by POX1 gene catalyses the oxidation of acyl-CoA to trans-2-enoyl-CoA in a reaction which is coupled with the one performed by the catalase Cta1, due to the release of H_2O_2 in the first reaction. The next step is the conversion of trans-2-enoyl-CoA to 3-ketoacyl-CoA thanks to the 3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase activities of the Fox2 enzyme. Finally, the gene product of *POT1* catalyses the cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA (Ploier, Daum and Petrovič, 2014).

Nutrient transport and sensing lead to the activation of signalling pathways

Glucose

Glucose is imported into the cell via different membrane proteins that allow a facilitated distribution of the nutrient. These proteins are called Hxt (hexose transporters).

Yeast possesses 20 *HXT* genes that belong to the family of hexose transporters (in particular, *HXT1-17, GAL2, SNF3* and *RGT2*), whose function is redundant, as shown by the fact that deletion of each single gene does not prevent growth on glucose.

Genes Hxt1-7 are the most important transporters and deletion of all 7 genes prevents growth on glucose.

It is not surprising that yeast is endowed with a great number of glucose transporters and sensing mechanism, because, as we saw before, glucose is the fundamental carbon source for yeast metabolism. Moreover, *S. cerevisiae*, due to its lifestyle and habitat, is exposed to very different concentrations of glucose and other sugars and it needs to react accordingly for maintaining a good homeostasis.

Consistently, the different functional transporters are dedicated to transport of glucose under different concentrations and have different affinity for the sugar (1mM<Km<100mM) and their expression in strictly regulated by the growth condition (Busti *et al.*, 2010).

Even if they belong to the hexose transporter family, *SNF3* and *RGT2* gene products have lost the ability to transport the sugar into the cell. They act as sensor of extracellular glucose concentration and therefore allow the cell to adjust glucose uptake and metabolism. In particular, Snf3 is sensible to low levels of extracellular glucose, while Rgt2 senses high levels of the hexose (Kayikci and Nielsen, 2015). The action of these two sensors generates an intracellular signal, which regulates the transcription of HXT genes. This signal is mediated by the casein kinases encoded by *YCK1* and *YCK2* with regulation by PKA pathway, which leads to the expression of the corrects HXT genes.

Yeast also possesses a glucose-sensing system that relies on a G-protein coupled receptor (GPCR) that is able to activate Protein Kinase A (PKA) signalling pathway. This kind of receptor is usually connected to 3 intracellular protein subunits (α, β and γ) that, in response to the ligand interaction dissociate into a α monomer and a βγ dimer, activating the downstream signal.

In this particular system, the GPCR is the Gpr1 protein, while the α subunit is encoded by *GPA2.* Many candidate genes have been identified as putative β and γ subunits for

a complex with Gpa2: in particular the kelch proteins Gpb1 and Gpb2, but it was shown that, even if they have a role in PKA signalling they do not actually behave as β and γ proteins. Gpa2 has a GTPase activity stimulates the activity of adenylate cyclase Cyr1, which catalyse the reaction converting AMP to cAMP, activating PKA signalling (see p. 32).

Sensing intracellular glucose is as important as sensing sugars outside the cell and concur to activate some of the same pathways. One of the most relevant system is the Ras pathway. This pathway is composed by a small monomeric G-protein (Ras1 or Ras2), which is active when bound to GTP and inactive after GTP hydrolysis. Therefore, the proportion of GTP- and GDP-bound Ras must be finely regulated. This regulation is the result of the action of guanine nucleotide exchange factors (GEFs, encoded by *CDC25* and *SDC25*), and GTPase Activating Proteins (GAPs, encoded by *IRA1* and *IRA2*). GEFs mediate activation of Ras, while GAPs stimulate hydrolysis of GTP and inactivation of the G-protein (Busti *et al.*, 2010).

The molecular mechanism by which Ras proteins are activated by glucose are not completely clear. The main hypothesis involves an action of glucose-6-phosphate on Cdc25/Sdc25, since phosphorylation of glucose is required for this activation (Rødkaer *et al.*, 2014).

Another intracellular glucose signalling pathway is the one that relies on Hxk2 and Mig1 and that regulates the activity of the kinase homologous to the mammalian AMP-activated kinase, which in yeast is encoded by *SNF1* gene.

Hxk2 is a so-called “moonlighting protein”. Aside from its cytoplasmic function in glycolysis, it was shown that it also localizes to the nucleus, where it interacts with DNA and with the transcriptional factor Mig1, inhibiting the expression of several genes, like the invertase *SUC2* (Rødkaer *et al.*, 2014).

The Snf1 kinase, instead, sense when glucose levels are low. Unlike its mammalian counterpart, the activation is not due to AMP, but is correlated with the levels of ADP. It is therefore strictly linked to the energetic state of the cell. Anyway, the specific signal triggering this activation has not been identified (Kayikci and Nielsen, 2015).

Nitrogen

Yeast possesses several permeases dedicated to the transport of amino acids and ammonium (Figure 12). The variety of nitrogen sources that yeast can metabolize results in the need for specialized sensing mechanisms to discriminate the nitrogen containing compounds.

A big difference between sensing of extracellular nitrogen and glucose in yeast is that the former relies both on non-transporting sensors and transceptors, capable of importing the molecule while transmitting the signal into the cell.

Permease	Substrate(s)	Affinity
Agp1	Asparagine, glutamine, other amino acids	Low
Agp3	Glutamine	High
Alp1	Arginine	—[a]
Bap2	Leucine	High
Bap3	Cysteine, leucine, isoleucine, valine	High
Can1	Arginine	High
Dip5	Dicarboxylic amino acids	High
Gap1	L-Amino acids	High
Gnp1	Glutamine	High
Hip1	Histidine	High
Lyp1	Lysine	High
Mep1	Ammonium	High
Mep2	Ammonium	High
Mep3	Ammonium	Low
Mmp1	*S*-Methylmethionine	High
Mup1	Methionine	High
Mup3	Methionine	Low
Ort1	Ornithine	—
Put4	Proline	High
Sam3	*S*-Adenosylmethionine	High
Tat1	Valine, leucine, isoleucine, tyrosine, tryptophan, histidine	Low
Tat2	Tryptophan, tyrosine	High
Uga4	γ-Aminobutyrate	—
Yct1	Cysteine	High

[a]—, not reported.

Figure 12 - The different permeases present in yeast for amino acids and ammonium (Zhang et al.*, 2018)*

Extracellular amino acids are sensed by the SPS sensor system. In their presence, the signal originating from this sensor induces the expression of genes involved in amino acid transporters and metabolic enzymes.

One of the key components is the protein encoded by *SSY1*, which shows great similarity with the amino acid permeases, but does not transport them. The N-terminal region of Ssy1 interacts with Ptr3 and Ssy5, constituting the sensor. In the presence of amino acids that in a process that involves phosphorylation and ubiquitination, activate the downstream effectors Stp1 and Stp2 allowing them to enter into the nucleus, where they act in combination with the transcriptional activators Dal81-Uga35 (Rødkaer *et al.*, 2014).

In addition to sensing via the SPS system, yeast cells possess several permeases that are needed for import of amino acids into the cell and for their sensing. This kind of proteins are called transceptors (Van Dijck *et al.*, 2017).

One of the first identified transceptors is the Gap1 permease, which mediate the import into the cell of all the canonical L-amino acids, plus L-citrulline, 4-aminobutyric acid, some D-amino acids. Studies have shown that Gap1 is able to activate PKA signalling. This activation does not require catabolism of the amino acids and the transmission of the signal is not due to cAMP production. Ammonium addition also have the same effect of amino acids addition on PKA activation. This signal is mediated by the three amino acid transceptors encoded by *MEP1, MEP2*, with the bigger effect being due to Mep2 (Steyfkens *et al.*, 2018).

Intracellular nitrogen is sensed by a highly conserved pathway that rely on the Tor kinases and the complexes in which they participate. Contrary to the multicellular eukaryotes, yeast possesses 2 different Tor kinases, encoded by *TOR1* and *TOR2*. Their gene products belong to the phosphatidylinositol protein kinase family, and form two different complexes (composed as in Table 2), called TORC1 and TORC2.

TOR complex	Subunits
TORC1	Tor1 or Tor2 (mainly Tor1)
	Lst1
	Kog1
	Tco89
TORC2	Tor2
	Avo1
	Avo2
	Avo3
	Bit6

Table 2 - Composition of yeast TOR complexes 1 and 2

TORC2 regulates mainly actin cytoskeleton and cell polarity and it is activated by changes in plasma membrane tension (Eltschinger and Loewith, 2016), TORC1 is activated by intracellular amino acid availability and regulates processes related to cellular growth (Loewith, 2010).

The localization of TORC1 is the vacuolar membrane. There, it is activated by the Ego complex, composed of Ego1, Ego2, Ego3, Gtr1, and Gtr2 proteins. The Ego proteins are the structural subunits and function in recruiting the complex to the vacuolar membrane and stabilizing the interaction. Gtr1/2 are the regulatory subunits, which form a heterodimer, which functions as a GTPase and it is completely active when GTP is bound to Gtr1 and GDP is bound to Gtr2. Both GEFs (Vam6 for Gtr1) and GAPs (L-Leucyl-tRNA and SEACIT complex for Gtr1, Lst4-Lst7 and SEACAT complexes for Gtr2) are present in yeast cells and respond to amino acid presence (Figure 13). Moreover, free tRNAs have been shown to inhibit TORC1 independently of the Gtr1/2 heterodimer (Zhang *et al.*, 2018).

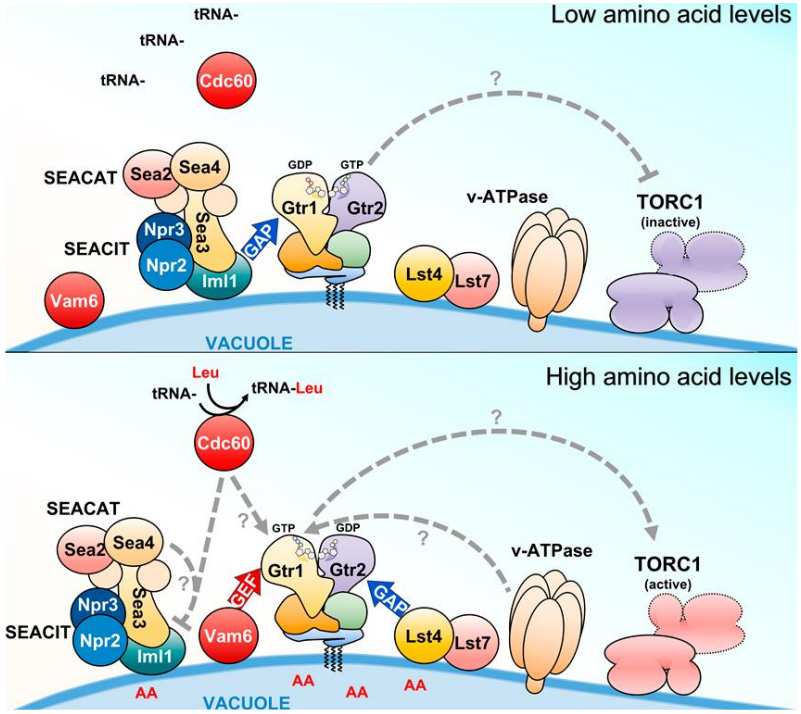


Figure 13 - Activation of TORC1 via the Gtr1-Gtr2 heterodimer (Nicastro et al., 2017)

Effect of the signal transduction pathways

In the last chapter, signals activating many signalling pathways have been mentioned. These pathways have pleiotropic effects on the cell growth, metabolism and energy and they are really important for the cell homeostasis. It is, therefore, important to understand what their main targets are and which cellular processes they affect.

PKA

The signal transduction pathway dependent on Protein Kinase A is the most important for glucose signalling, as demonstrated by the fact that 90% of yeast transcriptional changes due to a shift from a non-fermentable carbon source (like galactose) to glucose can be obtained as well by expression of a hyperactive Ras2 protein (Zaman *et al.*, 2009).

The purpose of PKA signalling in the presence of glucose is to positively regulate growth to allow the fast utilization of the preferred carbon source. Therefore, glycolysis is up-regulated, while genes required for utilization of non-fermentable carbon sources (respiration and gluconeogenesis) is repressed, as are the genes related to stress response (Busti *et al.*, 2010).

Glucose allow cells not only to grow faster but also to sustain a higher biosynthetic activity, which is consistent with the bigger cell size of glucose-grown cells. A way in which PKA signalling regulates this process is the induction of genes required for ribosome biogenesis. In particular, PKA regulates the localization of Sfp1 transcription factor, targeting it to the nucleus where it promotes transcription of ribosome biogenesis (Ribi) and ribosomal protein (RP) genes (Marion *et al.*, 2004).

PKA effect on the suppression of the stress response is mediated by action on the transcription factors encoded by *MSN2* and *MSN4*. This regulation is also at the level of the localization of the proteins (Görner *et al.*, 1998). Msn2/4 bind to the promoter of genes containing the STRE element. These genes participate in the response to a variety of different stresses regulating different processes like detoxification from reactive oxygen species, carbohydrate metabolism, proteolysis and protein folding (Gasch *et al.*, 2000).

Yak1 kinase is also a target of PKA action. It was first thought to function in a parallel pathway antagonizing PKA function. In particular, it phosphorylates and positively regulate Msn2 and Hsf1 proteins (Lee *et al.*, 2008). The latter is a transcription factor that express another category of stress-related genes, which contain the HSE element, involved in various stresses as well as in the diauxic shift. Moreover, its phosphorylation acts on the Crf1 transcriptional repressor antagonizing the transcription of ribosomal genes. Just like for the Msn proteins, PKA phosphorylation affects Yak1 localization and keeping it out of the nucleus, indirectly regulating the processes (Busti *et al.*, 2010).

When PKA is active, cells are in optimal growth conditions and do not need to stop. Therefore, the activity of a protein kinase like Rim15, whose role is mainly in governing the entry into stationary phase, must be stopped. Rim15 generally downregulates growth acting mainly by activating Gis1 (post diauxic shift transcription factor) (Pedruzzi *et al.*, 2000). Other targets of its regulation are Igo1 and Igo2 (Talarek *et al.*,

2010), which serve a protective role for the mRNA of the genes required for the entry into G_0 phase. As it will be shown by the following paragraphs, signalling pathways are not linear, but fully interconnected. Therefore, it is not surprising that Rim15 also acts on Msn2/4 and Hsf1 (Lee *et al.*, 2013). PKA phosphorylation of Rim15 inhibits its kinase activity (Reinders *et al.*, 1998).

PKA also regulates the activity of the Rgt1 transcription factor, leading to the expression of HXT genes (Busti *et al.*, 2010).

TORC1

As for PKA, a wide number of cellular processes are regulated by TORC1. The stimuli that activate these two pathways are different, but in both cases the meaning is that the cell is growing in a good environment with good nutrients. Therefore, it is not surprising that many downstream targets are common between the two kinase complexes.

TORC1 action is mainly mediated by two effectors: the kinase encoded by *SCH9* and the Tap42-protein phosphatase complex (Loewith, 2010).

Sch9 is a protein kinase finely regulated by TORC1 phosphorylation and, therefore, it is used as a readout for TORC1 activation (Urban *et al.*, 2007). Its targets include Rim15, which is kept in the cytoplasm, blocking its action. Moreover, Sch9 regulates ribosome biogenesis acting on RNA polymerases I, II and III (Smets *et al.*, 2010), autophagy and lifespan regulation (Sampaio-Marques *et al.*, 2011) and sphingolipid metabolism (Zhang *et al.*, 2018).

The other main branch of TORC1 signalling is due to its action on Tap42. TORC1-dependent phosphorylation promotes its association with the protein phosphatase 2A (PP2A) complex and the PP2A-like complex containing the Sit4 phosphatase. Tap42 interaction lead to inhibition pf the complexes. This result in a repression of nitrogen catabolite repression (NCR) and stress response, one of the main biological process regulated by PP2A and PP2A-like (Zhang *et al.*, 2018).

TORC1 also directly regulates Sfp1, as an additional control on ribosome biogenesis (Loewith, 2010).

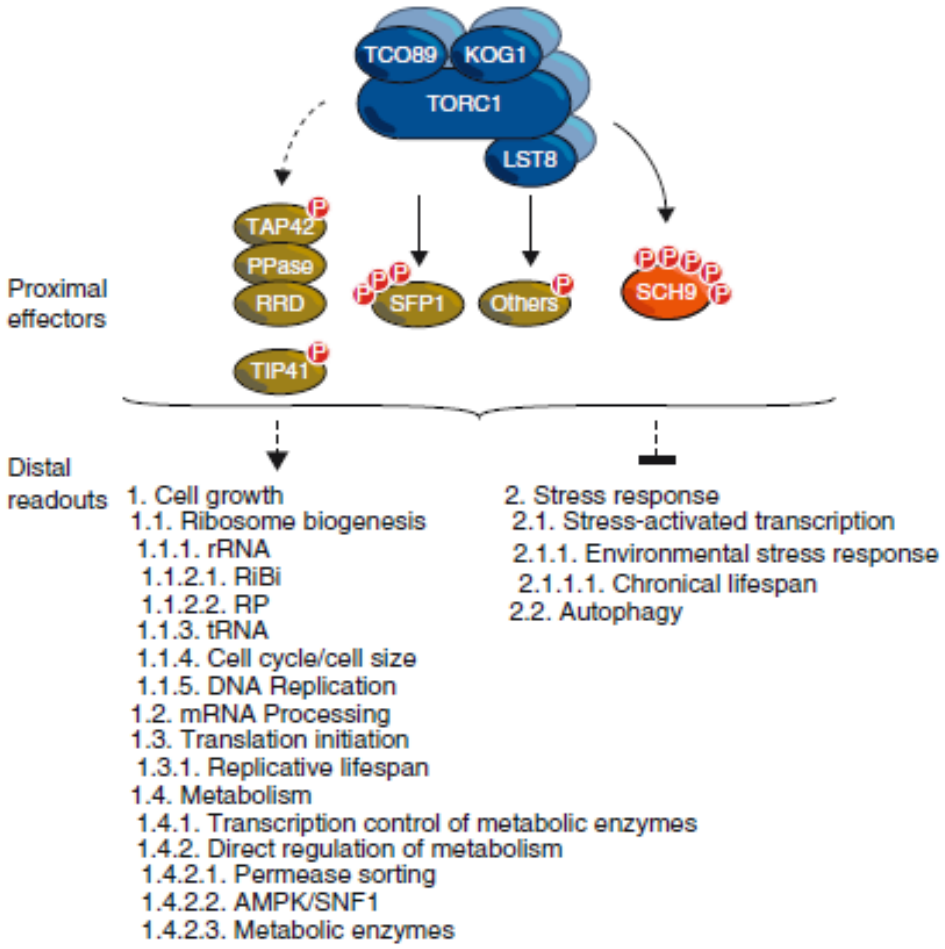


Figure 14 - TORC1 direct effectors and the pathways they regulate (Loewith, 2010)

Snf1

As already mentioned, the Snf1 kinase senses low levels of glucose and is therefore very important for the shift from a fermentative to a respiratory metabolism. A glucose level of 0.2% or below activates Snf1 by relieving its autoinhibition and promoting the binding of Snf4 regulatory subunit (Kayikci and Nielsen, 2015).

Once activated, Snf1 phosphorylate different targets having different transcriptional, post-transcriptional and post-translational effects. In particular, Mig1 phosphorylation by Snf1 causes its dissociation from DNA and allows the expression of the genes repressed by glucose (Kayikci and Nielsen, 2015). On the other hand, Snf1

phosphorylation of Cat8 induces the expression of gluconeogenetic genes required for growth on non-fermentable carbon sources, like *FBP1, MLS1* and *ICL1*. Phosphorylation of Adr1 induces the expression of genes required for ethanol and fatty acid metabolism (Young *et al.*, 2003). The latter is also affected at a post-translational level by an inhibitory phosphorylation of Acc1 and an activating modification of Fox2 (Usaite *et al.*, 2009). Moreover, phosphorylation of Gcn4 blocks the biosynthesis of amino acids (Zaman *et al.*, 2009). As a general result, expensive biosynthetic pathways are shut down, while reactions that produce energy are up-regulated.

Finally, Snf1 induce stress response by phosphorylating Msn2 (De Wever *et al.*, 2005).

PKA, TOR and Snf1 pathways are closely interconnected. Not only they share a large number of cellular targets (Figure 15), but the crosstalk between the pathways is relevant. For example, PKA and TORC1 negatively regulate Snf1 activity and localization, while Snf1 activity seems to inhibit the activation of Sch9 (Shashkova, Welkenhuysen and Hohmann, 2015).

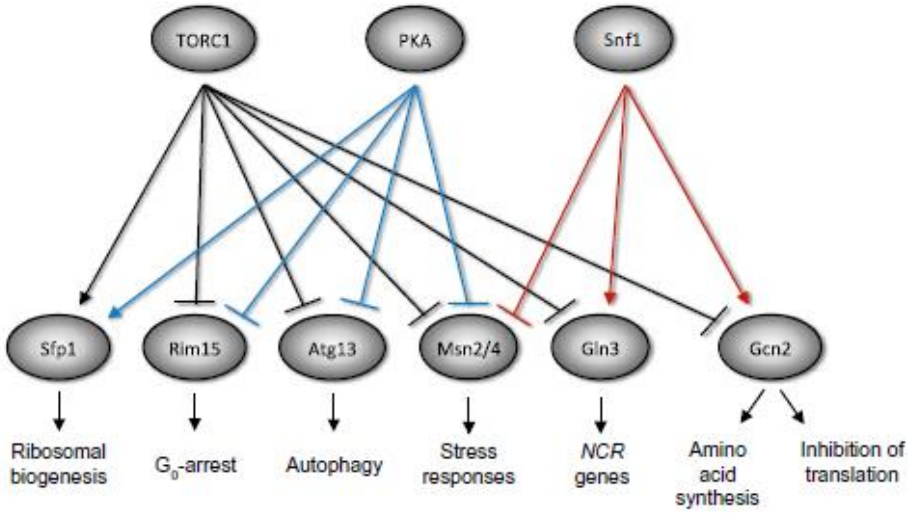


Figure 15 - Common targets of the signalling of TORC1, PKA and Snf1 pathways (Rødkaer et al., 2014)

Regulation of nitrogen transport and metabolism

Most of the work in this thesis concerns the utilization of nitrogen sources. Therefore, it is useful to spend a few words on the specific regulation of the cellular machinery required for the consumption of nitrogen sources.

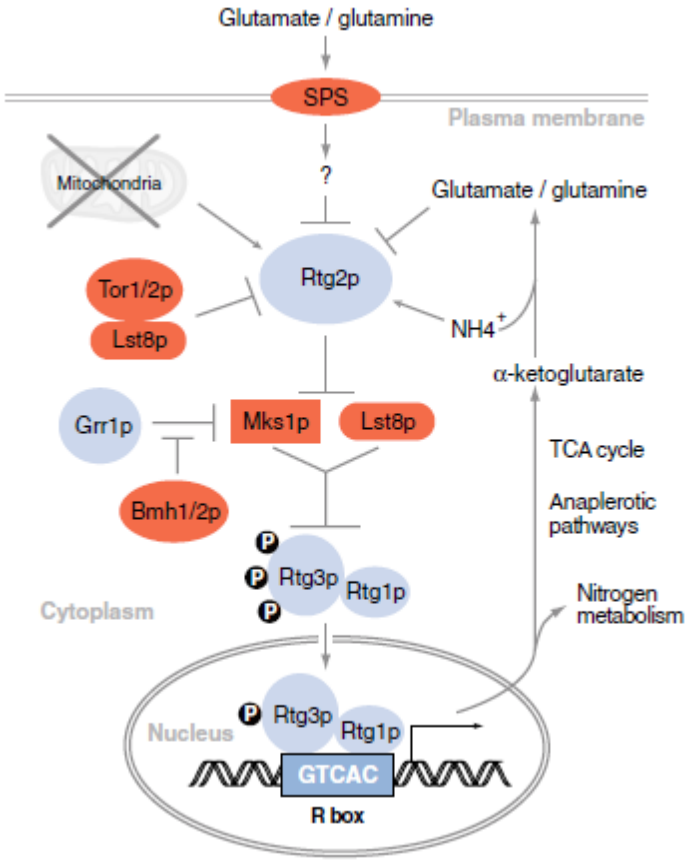


Figure 16 - Retrograde signalling in yeast (Liu and Butow, 2006b)

Yeast strains distinguish between preferred and not preferred nitrogen sources (green and red, respectively, in Figure 7). In the presence of a preferred nitrogen source, the NCR pathway is activated, even if the level of repression achieved is different for different preferred nitrogen sources (Godard *et al.*, 2007). This pathway is mediated by the action of 4 transcription factors: Gln3 and Gat1 (activators) and Gzf3 and Dal80 (repressors). Genes subjected to NCR have a GATA sequence in their promoter and these transcription factors bind to that. In the presence of a preferred nitrogen source Gln3 and Gat1 are phosphorylated by TOR and sequestered by the Ure2 protein and cannot activate the expression of the NCR genes. Moreover, Gat1 itself is under regulation by NCR (Zhang *et al.*, 2018). 41 genes have been identified as target of the NCR by classical genetics and molecular biology, and the list have been expanded with the contribution of transcriptomics analyses. They include broad-range and non-preferred amino acids and ammonium permeases, as well as many genes involved in

central nitrogen metabolism, including the "central hub" composed by *GLN1, GDH1-3* and *GLT1* (Godard *et al.*, 2007).

A pathway partly mentioned above is the general amino acid control (GAAC), which is activated during starvation for at least one amino acid. The outcome of this activated pathway is a down-regulation of protein synthesis via phosphorylation of the translation factor eIF2 (Conrad *et al.*, 2014) and an increase in amino acid biosynthetic processes, due to an increased activity of the transcription factor Gcn4, positively regulated in a transcriptional and translational way (Zhang *et al.*, 2018). As a rule of thumb, when the NCR is active, the GAAC is not and vice versa (Godard *et al.*, 2007).

As we have covered in the metabolism chapter, the TCA cycle is the point where carbon and nitrogen metabolism connect, due to the reaction producing α-ketoglutarate. The latter is under the control of the so-called retrograde signalling (RTG), which is a pathway by which the cell can sense mitochondrial disfunctions and which is very sensitive to the nitrogen source. As in the case of NCR, two transcriptional activators (Rtg1 and Rtg3) are responsible for the expression of RTG-regulated genes and are, in turn, subjected to regulation by TORC1 and SPS pathways, as well as other positive and negative regulators (summarized in Figure 16). Genes regulated by this pathway include those coding for enzyme in the first part if the TCA cycle (*PYC1, CIT1, ACO1, IDH1* and *IDH2*), the glyoxylate cycle citrate synthase Cit2 and Dld2, a mitochondrial enzyme that converts D-2-hydroxyglutarate to α-ketoglutarate (Liu and Butow, 2006a). Generally speaking, good nitrogen sources repress the expression of RTG-regulated genes. The strongest effect is, anyway, seen on glutamate, proline and citrulline (Godard *et al.*, 2007).

Studying modulation of growth by nutrients

Anybody who starts to work on the relationship between nutrients and growth, even in a tiny unicellular yeast, can't help but feel lost in all the metabolic and signalling pathways described so far. The first good help is always to look at omics data for our condition of interest: changes in transcriptome, proteome and metabolome are always useful to have a general picture of what is happening to that cell in that particular environment. This approach has been used to obtain some of the information contained in this introduction and also for the generation of the results of this thesis. Sometimes these analyses are not sufficient, and there are a few other approaches on which I would like to dwell upon.

Computational modelling

Classical genetics, molecular biology and biochemistry rely on a reductionist approach, that most of the time is not able to provide the full picture of a particular cellular process, let alone of the entire cellular physiology.

With the development of more powerful computers, it has become possible to perform simulations of biological systems that before were way too expensive from a computational point of view. This has led to the birth of a new science called systems biology. Systems biology is where biology, chemistry, computer science, mathematics and physics come together to solve complex problems.

The first step is always to define a mathematical model that is able to describe, in a formal way, the biological system of interest, which in our case will be the yeast metabolism. The model must be constructed in a way that allow for *in silico* predictions to be tested in the wet lab and to incorporate omics data into the simulation. The mathematical representation of a metabolic network is a graph, in which the metabolites are *nodes* and the reactions that connect them are *edges* (Cazzaniga *et al.*, 2014).

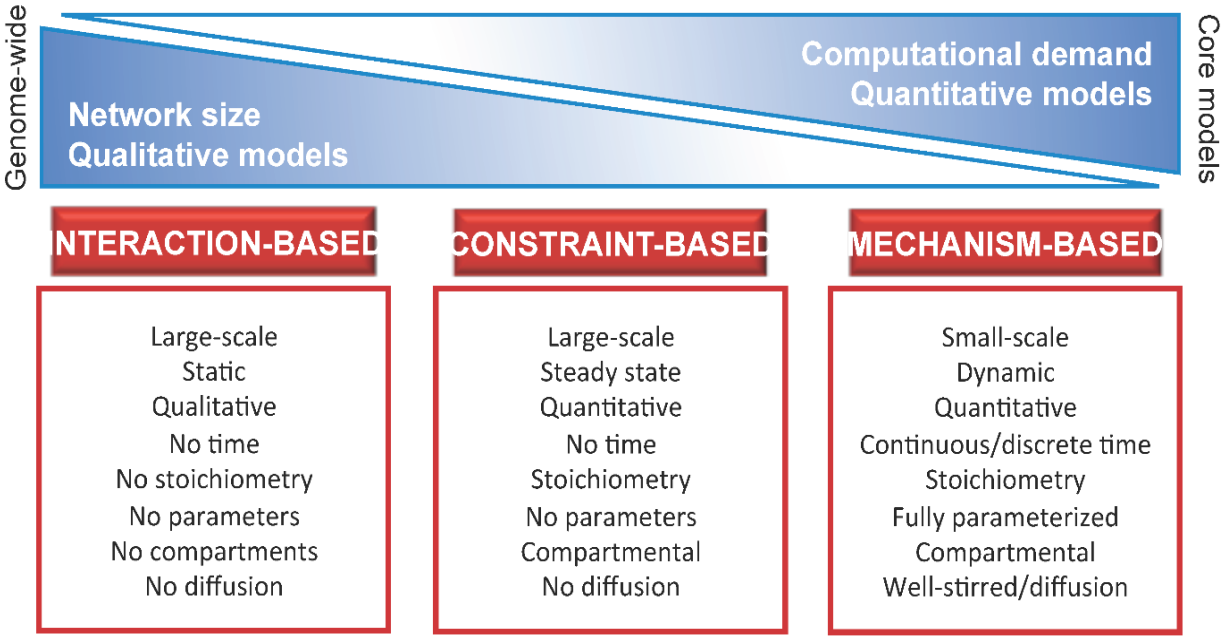


Figure 17 - Summary of the types of mathematical models in systems biology (Cazzaniga et al., 2014)

The types of models that can be used for describing a metabolic network can be classified in terms of scale as:

- Toy models, which comprise few reactions (e.g. model of glycolysis)
- Core models, that contain a higher number of reactions, usually involved in a particular sub-process (e.g. model of central carbon metabolism)

- Genome-wide models, that contain all the information known about metabolism and that have been possible after the genomic era and the development of high-throughput technology (e.g. YEAST7 (Sánchez *et al.*, 2017)). This model contains the higher number of information about the system and while this is advantageous, it has a high computational cost, which reflects on the kind of simulations performed.

A different, but related, level of classification regards the level of detail of the model (Figure 17).

Interaction-based models provide a qualitative description of the structure of the network being described. This kind of model usually starts from the information contained in pathway databases like KEGG (https://www.genome.jp/kegg/) and are useful to understand the presence of different functional modules in our biological systems. In other words, to have a topological view of the network (Stelling, 2004).

The description of the topology of the network is useful for identifying several of its emerging properties, and it is the first constraint that limits the functioning of the system. Other constraints are the stoichiometry of the reactions and their reversibility (Stelling, 2004). These can be used to construct the so-called constraint-based models, which consider the system under the assumption of a steady-state (therefore, they cannot take time into account) and can be used to compute all the possible physiological states the system can reach (in terms of distribution of the reaction fluxes), given the imposed constraints, which can also come from experimental data. The mathematical formalization of constraint-based models is the so-called stoichiometric matrix, in which the rows represent the metabolites and the columns represent the reaction (Cazzaniga *et al.*, 2014).

The dynamic simulations are the most fine-grained and the most predictive in systems biology. To perform those, there is a need for a lot of information about the system. In fact, in addition to the relationship between the components of the model, the kinetic parameters of the reactions and the starting conditions of the system must be known (Stelling, 2004). Because of that, even if these parameters are known, this kind of modelling requires specific the knowledge of specific mathematical formalisms (based on differential equations) and it has a high computational cost. For this reason, dynamic mechanism-based approaches are mostly used for toy or core models.

Flux Balance Analysis

Throughout the chapter of this thesis a particular kind of constraint-based model will be used, which is called flux balance analysis (FBA). FBA was not born within systems

biology, but instead it was borrowed from metabolic engineering, where it was used for the enhancement of production. Due to its initial purpose, FBA does not simply show every possible flux distribution given the constraints, but instead select the top-scoring distribution according to an objective. The objective function is selected by the user according to the scientific problem to be solved (e.g. maximization of biomass or metabolite production) (Feist and Palsson, 2010). This assumption is very useful from a production point of view. In regard to the maximization of biomass, cells are undoubtedly more complicated than that, as we have discussed in the previous chapters. Nevertheless, especially for the evolution of microorganisms, there is a very high selective pressure to have an optimal growth in a given environment. For this reason, keeping in mind that cells do not necessarily need to accumulate the maximum amount of biomass possible in every occasion, this objective function can be a useful approximation, and has been used to predict the behaviour of different microbial species (Burgard and Maranas, 2003).

In addition, many other objective functions can be implemented, depending on the biological process of interest. For example, maximizing the ATP production is a useful approach to study mitochondrial metabolism (Cazzaniga *et al.*, 2014).

Directed evolution to monitor adaptation

A different approach to study how yeast or any other microorganism reacts to nutrients is to observe the evolution of the strain under exam when it is put in an environment which is different from its ideal habitat.

One of the typical conditions to which yeast cells need to adapt is a medium with a poor carbon source or with a low amount of it. Yeast strains are cultivated for several generations in that particular conditions and if the right mutation arises that is able to allow for a better growth, it should be possible to identify by sequencing a pathway or a cellular function of interest which is modified and give the cells their phenotype.

Evolution experiments can be performed either from a population starting from a single clone, or more initial strains can be crossed to have more starting variability, and this can influence the recovered mutations (Figure 18).

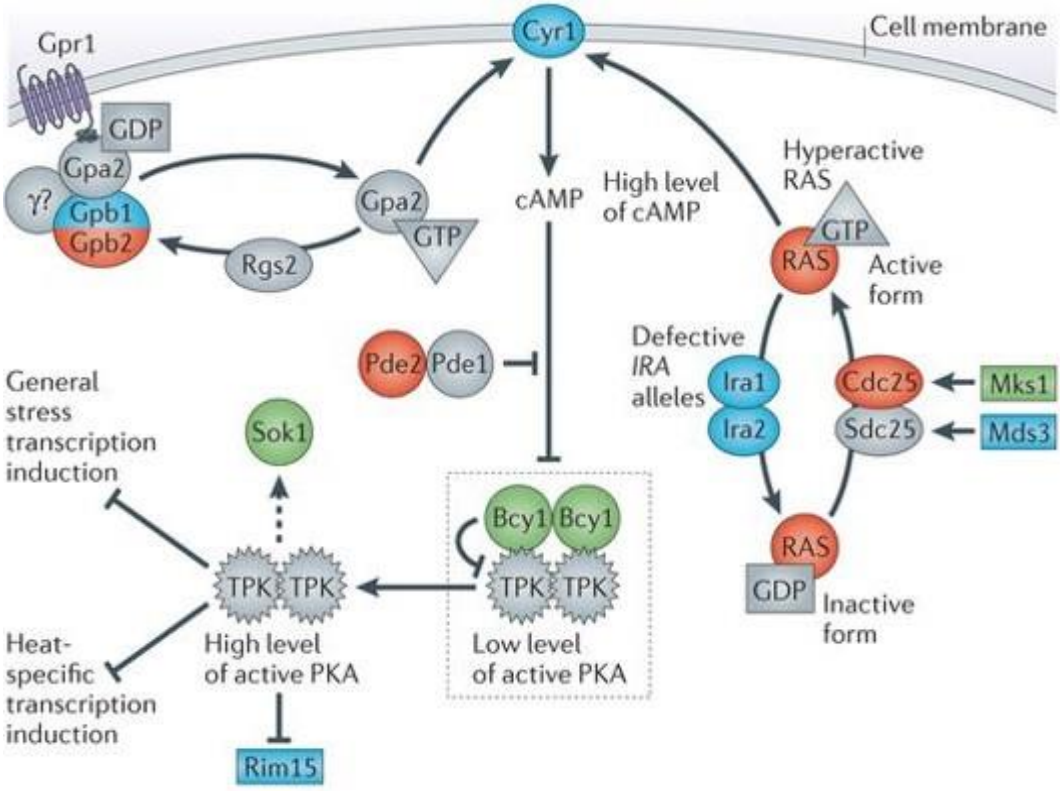


Figure 18 - The Ras pathway is the target of multiple mutation in yeast adapted to unfavourable conditions. Genes in red are found mutated in populations starting from a single clone, those in green in crossed population and blue in both (Long et al.*, 2015)*

From these experiments, it is easy to expect the emergence of mutations in genes involved in the specific metabolic pathway for the consumption of the nutrients present in the medium. Sometimes that happens, but mutations in the signal transduction pathways are also very frequent.

Glucose signalling, and in particular Ras/PKA/cAMP pathway, is one of the most affected from these mutations, as can be observed in Figure 18 (Long *et al.*, 2015).

To sum up, directed evolution can be a useful way both to elucidate unclear metabolic mechanisms and regulation and to obtain useful mutant for various applications. What it important to keep in mind is that evolution already shaped genomes to exploit the environment for the best. As a result, it is not surprising that many of the yeast strains adapted to survive and proliferate in difficult conditions, display a disadvantage, compared to their reference background, when they grow in optimal media (Wenger *et al.*, 2011).

Chapter 1: Glutamate, a multi-purpose nutrient sustaining enhanced yeast growth

Brambilla L[1,4], Gnugnoli M[1,4], Damiani C[2,4], Colombo R[2,4], Pescini D[3,4], Porro D[1,4], Frascotti G[1,4], Nicastro R[1,4], Airoldi C[1], Gotti L[1,4], Alberghina L[1,4], Vanoni M[1,4]

1. Department of Biotechnology and Biosciences, University of Milano-Bicocca
2. Department of Informatics, Systems and Communication, University of Milano-Bicocca
3. Department of Statistics and Quantitative Methods, University of Milano-Bicocca
4. SYSBIO.it - Centre of Systems Biology, Milan

Abstract

We investigated growth, metabolism and transcriptional profiles of yeast cells grown in minimal media supplemented with either ammonium or glutamate, both considered good nitrogen sources. During the exponential phase, cells using glutamate as a nitrogen source have a larger cell size compared to cells grown in presence of ammonium. In stationary phase, glutamate-supplemented accumulate higher biomass levels. We refer to this behaviour as “enhanced growth”.

In glutamate-grown cells, the deepest transcriptional and metabolic rearrangement takes place after glucose exhaustion and show a profound alteration of the metabolism of storage molecules during growth on glutamate with an accumulation of trehalose and fatty acids, which correlate with a higher stress resistance.

Flux balance analysis simulations with a core model of yeast metabolism correctly predicted the optimal growth yield and computational analysis of the flux distribution identified a different allocation of oxygen as responsible for the observed interplay between ethanol and glutamate.

Introduction

Every living organism needs to tightly regulate its growth depending on the environment that surrounds it. This is true from the single-celled microorganisms to the mammals. Failure to correctly respond to environmental changes can be dangerous for the correct homeostasis of the cell. For this reason, the molecular basis and the regulation of many signalling pathways are conserved from bacteria to man (Smets *et al.*, 2010).

Mammalian cells are part of a greater organism, and their growth need to be tightly regulated both by nutrient availability and growth factors. Cells that do not regulate their growth correctly can give rise to pathologies like cancer. This is true on many cellular levels, from the senescence to the regulation of metabolism. In fact, a deregulated metabolism is considered a hallmark of cancer (Hanahan and Weinberg, 2011).

The yeast *Saccharomyces cerevisiae*, on the other hand, primarily regulate its growth in response to changes in the availability of nutrients in the environment and to the presence of stresses. Evolution selected a variety of sensing mechanisms and signal transduction pathways, that enable yeast to apply different strategies to react to environmental cues. One of the main examples is the so-called “Crabtree effect”, in which *S. cerevisiae* can ferment glucose to ethanol even in the presence of oxygen if the sugar concentration is high enough. This allows yeast to grow faster and outgrow competing microorganism, whose growth is also inhibited by the ethanol produced(Broach, 2012).

It is also important to keep in mind that high growth rate is not necessarily the goal of an efficient nutritional regulation. It was shown that there is actually a trade-off between growth rate and biomass yield (Bachmann *et al.*, 2013), which can be more or less important depending on the growth condition (Wortel *et al.*, 2018). Moreover, not all the nutrients are used to sustain growth and cellular components, but some are stored as reserve compounds, like lipids and storage carbohydrates. In yeast, glycogen is synthesized during the exponential growth on glucose, and serves mainly as stored carbon and energy, while trehalose is accumulated during the post-diauxic growth on ethanol and acts as a stress protectant (Francois and Parrou, 2001).

One of the fundamental nutrients that yeast needs to react to is nitrogen. *S. cerevisiae* is able to grow on a variety of nitrogen sources that and shows different responses based on the amount and quality of the nitrogen source. Nitrogen-containing compounds can be classified into preferred and poor nitrogen sources for a given yeast strain. This classification can depend either on how well they are able to sustain its

growth or by the ability of the preferred source to cause the repression of genes required for the utilization of alternative nitrogen sources (Godard *et al.*, 2007).

There is not much work investigating the influence of different nitrogen sources on the physiology of yeast during the whole growth, and even less that focuses on changes in the transcriptome and metabolome that result from the interaction between the growth phase and the nitrogen source.

Glutamate is a very important amino acid that can be utilized by yeast cells as a nitrogen source. It is, together with glutamine, one of the main compounds used as nitrogen donors for yeast biosynthetic reactions (Cooper, 1982). The degradation and assimilation of glutamate is mainly due to the NAD-dependent oxidative deamination catalysed by the *GDH2* enzyme (Miller and Magasanik, 1990). This reaction yields ammonia and α-ketoglutarate, which is an important intermediate in the citric acid cycle. This fact place glutamate at the interface, not only between nitrogen biosynthesis and degradation of nitrogen containing compounds, but also between carbon and nitrogen metabolism. Ammonium sulphate is the standard nitrogen source for yeast synthetic media. For this reason, we used it as a reference to test how a combined nutritional source like glutamate is able to change yeast behaviour. In particular, we monitored the physiology of our strain in the two different conditions (always in the presence of glucose) and we performed transcriptomics and metabolomics experiments, followed by molecular and biochemical validations.

The following results describe the growth phase-dependent phenotype of the glutamate-grown cells, which we called “enhanced growth”. We investigated this phenotype with a combination of classic cellular biology, transcriptomics, metabolomics and computational modelling. We found a great alteration of transcription and metabolism in cells grown on glutamate, compared to those in ammonium. This leads to a higher biomass production, which we show to be the result of an interplay between carbon and nitrogen source, and to an accumulation of reserve carbohydrate and an enhanced stress resistance.

Materials and methods

Strains and Media

The prototrophic GRF18 (GRFc) strain (Brambilla *et al.*, 1999) was used in this study. Inoculum cultures were prepared from −80°C glycerol stocks on YPD agar plates containing 10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose and 20 g/L agar.

Media were prepared according to (Verduyn *et al.*, 1992) with 20 g/L glucose or 20 g/L ethanol as carbon sources and either 5 g/L ammonium sulphate or 14 g/L glutamate monosodium salt monohydrate as nitrogen sources.

Shake flask cultures

Precultured cells were inoculated in 500 mL Erlenmeyer flasks containing 100 ml of minimal medium (Verduyn *et al.*, 1992) with the same composition of cultures. The starting OD_{660}, ranging between 0.02 and 0.2, was suitable to reach the desired growth phase the day of sampling.

The cultures were grown at 30°C in a rotary shaker at 150 rpm.

Inhibition of respiration was obtained with the addition of Antimycin A (Sigma-Aldrich) at a final concentration of 5 µM to exponentially growing cells.

To test the effect of the inhibition of fatty acids synthesis, a sublethal dose (1 µM) of cerulenin (BioVision, Inc.) was added to the cultures at the beginning of the exponential growth phase ($2\text{-}3 \cdot 10^6$ cells/mL)

To monitor the recovery from oxidative stress H_2O_2 0.5 mM was added to exponentially growing cells (OD_{660}= 0.5).

Batch cultivation in fermenter

Cultivations in fermenter were run in Biostat-B fermentors (B-Braun), filled with 1.5 L of defined medium with vitamins and trace metals (Verduyn *et al.*, 1992). The initial glucose concentration was 20 g/L, nitrogen sources were 5 g/L ammonium sulphate or 14 g/L glutamate monosodium salt monohydrate.

Cultures were grown at 30° C at pH 5.0 and 0.2 M KOH was used as a titrant.

For anaerobic experiments, glucose concentration was 50 g/l and the medium was supplemented with 10 mg/L ergosterol and 420 mg/L Tween 80 (Brambilla *et al.*, 1999). To maintain anaerobic condition the air supply was replaced by nitrogen gas (flow rate, 0.15 l/min, 0.1 v/v/m).

Glucose-limited aerobic chemostat cultivations

Chemostat cultivation was performed in Biostat-B fermentors (B-Braun). A defined medium with vitamins and trace metals was used (Verduyn *et al.*, 1992). The glucose concentration in the reservoir medium was 8 g/L. A constant working volume of 1300 mL was maintained via an effluent line coupled to a peristaltic pump. A dissolved oxygen concentration above 30% of air saturation was maintained by an air flow of 1.3/min (1 v/v/m) and a stirrer speed of 1000 rpm. The temperature was maintained at 30°C and the culture pH was kept at 5.0 by automatic addition of 2 M KOH.

Cultures were assumed to be in steady state when at least six volume changes had passed since the last change in growth conditions and the culture did not exhibit metabolic oscillations.

Measurement of growth parameters

Optical density of the culture was measured at 660 nm using an Ultrospec 500 pro spectrophotometer (Amersham).

Cell titre and volumes were obtained after 30s of sample sonication and appropriate dilution in a NaCl 0.9% solution, with a Coulter Counter model Z2 (Beckman Coulter) equipped with a 70 µm aperture tube.

Budding Index was calculated by scoring at least 300 cells under an Olympus CH20 microscope

Determination of dry cell weight

The dry cell weight (DCW) was determined by filtering 10 mL of culture broth through pre-dried 0.22 µm membranes filters. The filters were washed with demineralized water and dried to constant weight in a microwave oven (van Dijken JP *et al.*, 2000).

Determination of RNA, protein and neutral lipids contents by flow cytometry

At least 2·107 cells from growing cultures were centrifuged and fixed in 70% ethanol to allow the staining of intracellular components. To stain RNA, cells were washed with ice cold PBS (3.3mM NaH2PO4, 6.7mM Na2HPO4, 127mM NaCl. 0.2mM EDTA, pH 7.2), resuspended in 1 mL of propidium iodide staining solution (0.046mM propidium iodide in 0.05M Tris–HCl, pH=7.7; 15mM MgCl2) and incubated for 30 min in ice, protected from the light. Cell suspensions were separated by sonication for 30 s before being analysed. For protein staining, cells were resuspended in a fluorescein isothiocyanate solution (50 µg/mL FITC in 0.5MNaHCO3) for 30 min in ice and in the dark; cells were washed three times with PBS and sonicated before the analysis. For neutral lipids quantification, cells were stained with Nile Red solution (10 µg/mL Nile Red in PBS) for 10 min at RT; cells were then washed once in PBS and sonicated before the analysis.

Stained cells were analysed with a FACSCalibur (BD Biosciences) flow cytometer and excited at 488 nm with an Ion-Argon laser. 30000 cells were analysed for each sample.

Fluorescence microscopy of lipid droplets

To observe lipid droplets directly cells were washed with PBS and stained with BODIPY 493/503, added at a final concentration of 1 µg/mL and incubated for 10 min at RT in the dark. Cells were then observed under a Nikon Eclipse E600 fluorescence microscope, equipped with a 100X, 1.4 oil Plan-Apochromat objective using a standard fluorescein filter. Image acquisition was performed with a Leica DC 350F camera and the obtained files were processed with ImageJ (http://imagej.nih.gov/ij/).

RNA preparation and hybridization

Cells were collected by filtration at the indicated time points and total RNA was extracted with the phenol-chloroform method (Schmitt, Brown and Trumpower, 1990) from biological triplicate samples in the and analysed using Affymetrix Genechip Yeast Genome 2.0 Arrays, containing approximately 5,744 probe sets for 5,841 of the 5,845 genes present in *S. cerevisiae*. Messenger RNA was amplified and hybridized onto the array as recommended by Affymetrix.

Microarray data analysis

Raw data (CEL files) were normalized with the Robust Multi-Array method using the R package *simpleaffy*. Normalized data matrix values are expressed in log2 scale.

To evaluate the differential gene expression linear regression models and empirical Bayes moderated F statistics were used, as implemented in the R/Bioconductor package *Limma* (Ritchie *et al.*, 2015).

Functional analysis of DEGs

To identify Gene Ontology and KEGG terms over-represented in the differentially expressed genes identified by the transcriptomic analysis the online tool FIDEA (http://circe.med.uniroma1.it/fidea/) was used to perform an enrichment analysis. The p-value threshold was set at 0.05.

Refinement of the GO terms list obtained from FIDEA to remove redundant terms was performed with REVIGO (http://revigo.irb.hr/).

To obtain further information on the regulatory network of the DEGs, the “Rank by TF” tool of the Yeastract database (http://www.yeastract.com/) was used to predict a ranking of the main transcription factors (TFs) for responsible for the regulation of our genes of interest.

To investigate the interactions occurring between the gene products of some of the DEGs, the tool STRING (https://string-db.org/) was used.

Quantitative Real-time PCR

Reverse transcription of the extracted total RNA was done using the SensiFAST™ cDNA Synthesis Kit (Bioline). Quantitative Real-time PCR reactions were performed in a MiniOpticon system (BIO-RAD) with the SsoFast EvaGreen Supermix (BIO-RAD). Primer sequences are available on request. The CFX Manager software (BIO-RAD) was used to analyse the obtained data, which were normalized to the expression levels of the TAF10 housekeeping gene within the same sample.

Analysis of intracellular metabolites

Around 5 mg of exponentially growing cells (OD_{660}=0.7) were collected by filtration and quenched with 1.5 mL 50% MeOH at T<-40 °C. Samples were then centrifuged at maximum speed for 1 min and the supernatant was discarded. For metabolites extraction, 400 µl of ice-cold chloroform, 800 µl of 50% MeOH and 20 µl of 2 mM norvaline (as internal standard) were added and samples were mixed by vortexing for 30 min at 4 °C. Following centrifugation for 5 min at 4 °C, the resultant supernatant was concentrated through evaporation. MSTFA (N-Methyl-N-(trimethylsilyl) trifluoroacetamide) was used to derivatize samples in an automated WorkBench (Agilent Technologies). The analysis was carried out with 7200 accurate-mass Q-TOF GC/MS (Agilent Technologies). Data handling and analysis was performed with Mass Hunter and Mass Profiler Professional software (Agilent Technologies). Raw data were normalized on the internal standard (norvaline) signal and on the DCW of the sample.

Intracellular trehalose was extracted according to (Parrou and François, 1997). Trehalose concentration was determined using a Trehalose kit (Megazyme International Ireland).

Analysis of extracellular metabolites

2 mL of culture broths were taken from the shake flask, chemostat or batch cultures. Samples were centrifuged for 2 min at maximum speed in microcentrifuge. Supernatants were stored at −20°C until analysis.

Extracellular glucose and ethanol concentrations were determined by D-glucose–Hk and Ethanol enzymatic assays (Megazyme International Ireland), respectively.

Glutamate and AKG concentrations were determined by H-NMR analysis. Briefly, the medium pH was adjusted to 7.5, 750 µL of sample were frozen in liquid nitrogen and lyophilised. Samples were then resuspended in D_2O and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was added at a final concentration of 1.5 mM (as an internal control).

550 µL of samples were used for NMR spectrum acquisition with a Bruker Advance III 600 MHz spectrometer. Spectra were analysed with the MestReNova software.

Determination of glutamate metabolic fate

From a standard preculture cells were inoculated in a medium containing ^{13}C Glutamate (Cambridge Isotope Laboratories) at the usual concentration of 14 g/L. Samples were collected for the following 72 hours and both supernatants (as described above) and cellular extracts (extracted using the boiling ethanol method) (Gonzalez, Francois and Renaud, 1997) were analysed via NMR for the presence of ^{13}C compounds. In particular, samples were resuspended in H_2O:D_2O 9:1. Acquisition of the NMR spectrum in the absence of DSS allows to monitor the ^{13}C-labelled fraction of each compound under exam, thanks to the splitting of ^{1}H signals due to coupling with ^{13}C carbons. Then, after addition of DSS 2mM to the samples, the absolute concentration of each metabolite can be calculated.

Determination of enzymatic activities

Cellular soluble protein fraction was extracted as described previously (Landi *et al.*, 2015). Briefly, cells were centrifuged at 2000 g and washed twice with 20 mM phosphate buffer (pH 7.0); the pellets were resuspended in a buffer containing 1 mM EDTA and 0.1 mM PMSF and mixed with an equal volume of acid-washed glass beads. Suspensions were subjected to 5-8 cycles of vortexing, (shaking for 1 min and placing the sample on ice for another 1 min, alternatively), and the cell debris was removed by centrifugation at 13,000 g for 10 min at 4 °C. The total protein in the supernatants was determined by Coomassie Plus Assay kit (Thermo scientific) with bovine serum albumin as a standard, and enzyme activities were assayed at 25 °C using an Ultrospec 500 pro spectrophotometer (Amersham).

To determine catalase activity, 5.0–50-µl aliquots of soluble protein extract were added to 1.0 ml of 13 mM H_2O_2 in 50 mM phosphate buffer (pH 7.0). H_2O_2 decomposition was monitored at 240 nm (ε_{240}=43.6 M^{-1} cm^{-1}). One Unit of catalase activity is the amount of enzyme that catalysed the degradation of 1 µmol of H_2O_2/min.

Pck1and Mls (and Icl1) activities were determined according to (de Jong-Gubbels *et al.*, 1995).

3-Hydroxyacyl-CoA dehydrogenase activity was determined according to (Ueda *et al.*, 1985).

Determination of stress resistance

Exponentially growing cells were collected at an optical density between 0.3-0.5, concentrated and exposed to the following stresses: 49°C (10 min); sorbitol 4M (1 h);

LiCl 2M (3h); H_2O_2 2mM (1h). Serial dilutions of the suspension were plated on YPD plates and incubated for 3 days at 30°C.

Determination of cell viability

The percentage of dead cells within the population was determined by flow cytometry after staining with 4.6µM propidium iodide.

YENIM1 network reconstruction

The model has been extracted from YEAST 7 and is designed to evaluate the contribution of glucose, glutamate and ammonium/ammonia (only NH_3 is modelled as a simplification) to cellular growth. Mitochondrion and cytosol are the considered as compartments. Linear pathways have been lumped in a unique reaction. The model is structurally free from thermodynamically infeasible loops. The model includes about 100 reactions and 80 metabolites pertaining to the following pathways: glycolysis, TCA cycle, the mitochondrial electron transfer chain, GABA, glutathione, synthesis of palmitate and main amino acids (glutamate, proline, alanine, arginine, aspartate), urea cycle, glyoxylate, ethanol and acetate fermentation, pentose phosphate, trehalose, glycerol. The model includes a biomass formation reaction formulated according to biomass macromolecular composition estimations in (Cortassa, Aon and Aon, 1995).

Flux Balance Analysis

To run FBA simulations a stoichiometric matrix S and a set of constraints (upper and lower bound of fluxes) are required. The steady state constraint is defined by the equation $dx/dt = S \cdot v = 0$, where dx/dt are time derivatives of metabolite concentrations represented by the product of the $m \times n$ matrix S times the vector of fluxes $v = (v_1, v_2, \ldots, v_n)$, where v_i is the flux of reaction i, n is the number of reactions, and m is the number of metabolites. The ensemble of functional states that the system can reach given a boundary condition I determines the feasible solutions space $\Phi = \Sigma \cap I$. By exploiting linear programming, FBA allows for optimization of the flux through a weighted sum of fluxes. In particular, we used the COBRA Toolbox and the GLPK solver. Because FBA only returns a single solution, we exploited Flux Variability Analysis to assess the flux variation range consistent with the maintenance of the maximal growth obtained with FBA (Damiani *et al.*, 2017).

Single reaction deletion experiment

Each single reaction in the model is deleted by setting the boundaries (both the upper and the lower bound) of the allowed flux trough that reaction to 0. After the deletion FBA is newly performed to optimize biomass. The percentage ratio of the newly obtained biomass value over the baseline value (value obtained when flux is allowed

through all model reactions) is registered. Once the boundaries of the deleted reactions are set back to their original values a new single deletion is performed.

Sampling in the region of feasible solutions

As a complement to classic FBA, Monte-Carlo approaches are emerging with the aim of exploring the entire region of feasible flux distributions. In this work, we apply the simplex method with a random set of objective functions to be maximized, as in (Damiani *et al.*, 2017). The maximization of each of these objective functions gives a corner in the space of solutions.

Identification of the fluxes that are statistically different between growth conditions

To determine the fluxes that were most sensitive to changes in nutrient availability, we sampled the solution space of the model at control (e.g ammonium) and at a condition of interest (e.g. glutamate), exploiting the optimization of a random set of multi-weighted objective functions. For each flux, we then determined a Z-score that quantifies its significance of change and that can be positive or negative. *Z* is computed as $Z = (\bar{X}_1 - \bar{X}_2)/\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}$, where $\bar{X}_1$ is the average of the flux in the control, $\bar{X}_2$ is the average of the flux in the condition of interest, σ_1^2 is the variance for condition 1, σ_2^2 is the variance for condition 2, and *n* is the sample size.

Results

Physiological response to glutamate is affected by the diauxic growth phase

We compared growth-related physiological parameters of cells supplemented with either ammonium sulphate or glutamate as a sole nitrogen source during diauxic growth in shake flasks. Phase I, II and III are denoted by a white, light grey and dark grey colour throughout the paper. The sampling points are indicated as T1, T2 and T3 respectively. A fourth time point (T_{EtOH}), used in some of the experiments refers to exponential growth on ethanol.

Although during exponential growth (phase I, white background) cells grow with the same growth rate in both media (Figure A and Table), glutamate-grown cells display a significant alteration in the coordination between cell growth and cell cycle, witnessed by the longer budded phase and increased cell volume. Increased cell volume reflects a true increase in cellular accumulation of biomass, as indicated by RNA and protein distributions (Figure D).

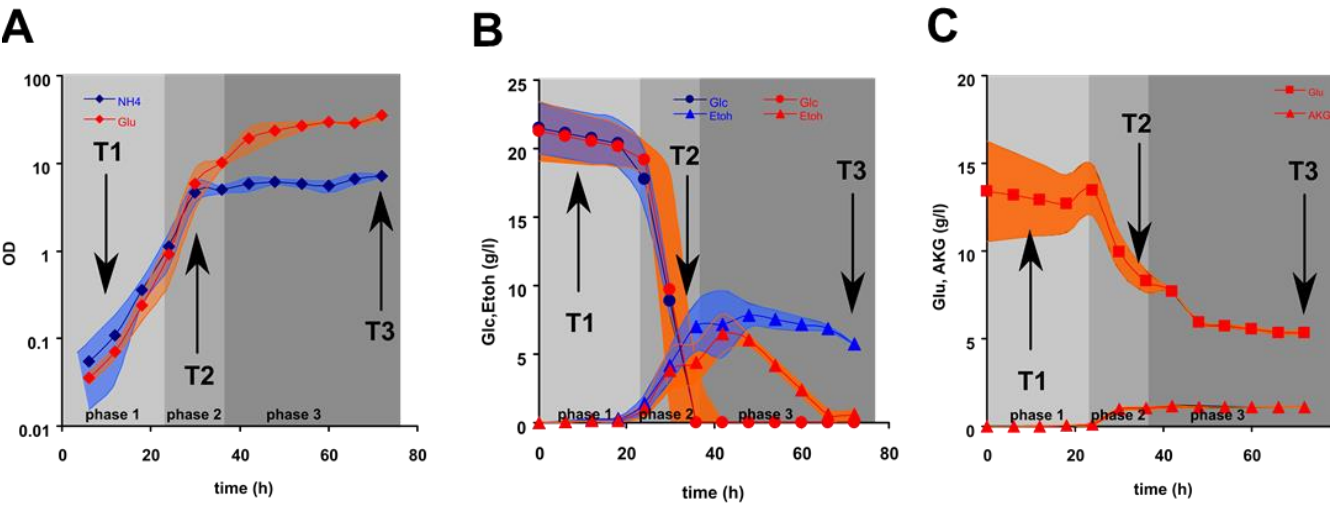
Glucose consumption and ethanol accumulation curves are very similar for both cultures, suggesting a similar distribution of the main metabolic routes (Figure B). Glutamate-grown cells release a large amount of alpha-ketoglutarate (AKG) in the growth medium. ^{13}C-labelled glutamate experiments show that almost 100% of AKG in the medium derived from glutamate (Supplementary table S2). In Phase I we observed a linear relation between consumed glutamate and excreted AKG, indicating that almost half of glutamate-deriving carbon moieties entering the cells during the fermentative growth phase are not assimilated (E).

Starting from Phase II the difference in cell size, RNA and protein content between cells grown in glutamate and ammonium-supplemented cells starts to attenuate (Table 1)

	T1		T2		T3	
	amm.	glut.	amm.	glut.	amm.	glut.
MDT (min)	151 ± 11.2	151 ± 24.8	N.A.	N.A.	N.A.	N.A.
TB(min)	106 ± 3.2	128 ± 1.6 (**)	N.A.	N.A.	N.A.	N.A.
Vol (fl)	42 ± 1	63 ± 2.4 (**)	43± 2	53.8 ± 2	48 ± 1	58.2 ± 1
Protein content (P)	216 ± 8	383 ± 12 (***)	317 ± 18	296 ± 2	359 ± 29	261± 8 (*)
RNA content (R)	190 ± 17	444 ± 31 (***)	180± 4	180 ± 81	230± 8	181 ± 81
RNA/P	1.18	0.97	0.86	1.00	0.90	0.92
Cells/ml	6.4E+06	6.2+E06	7.9E+07	1.8E+08	1.3E+08	7.6E+08
OD_{660}	0.3 ± 0.02	0.3 ± 0.08	7.1 ± 0.06	11.7 ± 0.90 (**)	8.2 ± 0.10	31.0 ± 1.15 (**)
DW (g/l)	0.12 ± 0.01	0.14 ± 0.01	1.7 ± 0.05	3.7 ± 0.25(**)	1.9 ± 0.07	8.9 ± 0.73(**)

Table 1 – Quantitative data about growth, size and biomass formation of ammonium- and glutamate-grown cells. Data are reported for T1, T2 and T3 time point, which correspond to those indicated in Figure A.

and the linear relation between glutamate consumption and AKG release is lost (Figure E). Concurrently, a positive bias in biomass formation becomes evident, either considering cell titre, optical density or dry weight, the latter being 3.5 times higher in glutamate compared to ammonium-grown cells (Table 1).



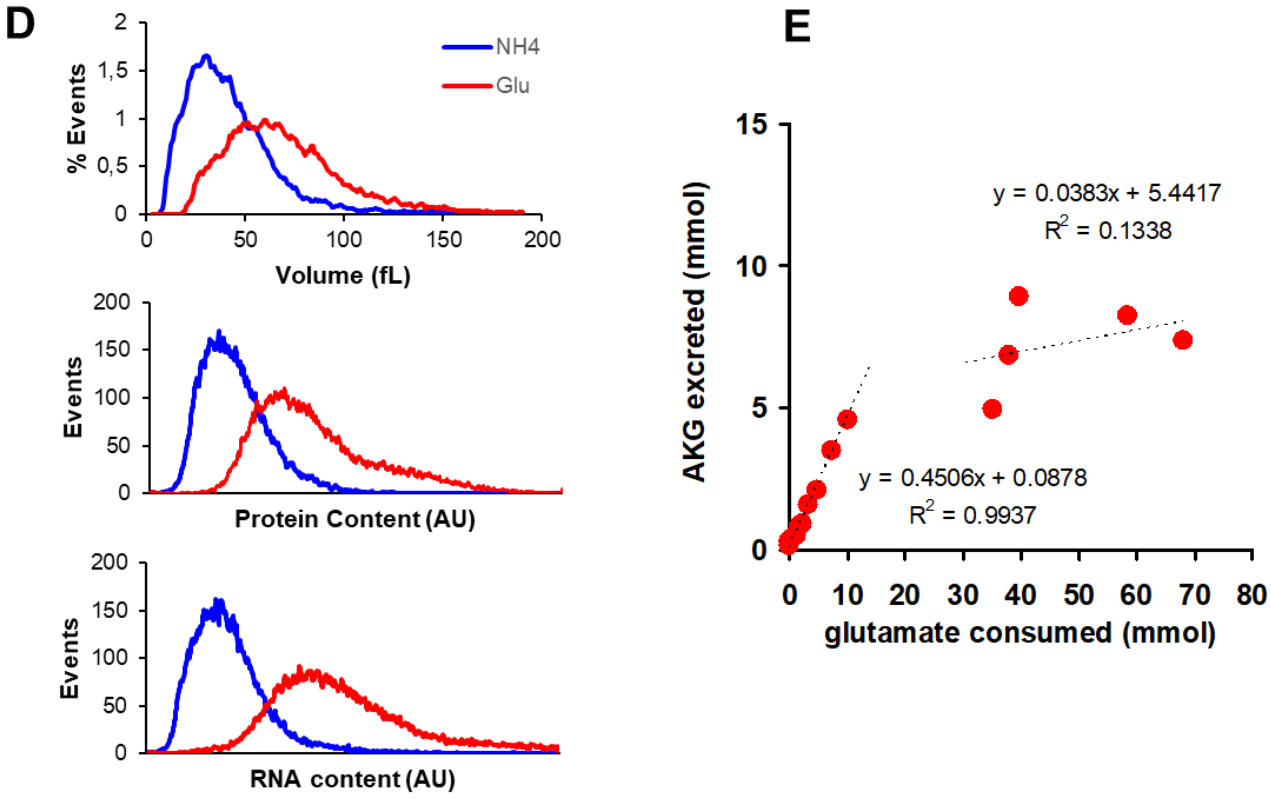


Figure 1 – (A) Glutamate induces visible effects on growth starting from glucose exhaustion. Cellular growth, expressed as culture optical density. Three main growth phases were defined in relation to presence of glucose and type of metabolism. For each phase, arrows indicate the chosen sampling point. Data represent an average growth curve, obtained by the interpolation of several independent experiments. For each point standard deviation is represented.
(B) Glucose and ethanol evolution during growth on ammonium sulphate (blue symbols) and glutamate (red symbols) medium. Graph shows the interpolation of data coming from several independent experiments. For each point standard deviation is represented. Other symbols like in A.
(C) Glutamate and 2-oxoglutarate (AKG) evolution during growth on glutamate medium. Graph shows the interpolation of data coming from several independent experiments. For each point standard deviation is represented. Other symbols like in A.
(D) Glutamate influences cellular dimension during exponential batch growth. Single cell protein and RNA distributions obtained with a cytofluorimetric analysis. Glutamate-grown cells (red line) display a relevant increase of their average values respect to ammonium (blue line).
(E) Sustained release of 2-oxoglutarate formed from glutamate. Linear relation between glutamate consumed and AKG secreted. Points were obtained in phase1 (exponential growth on glucose). For each sampling point the mmoles of glutamate consumed were plotted against the mmoles of AKG found in the medium. Dotted line represent a linear interpolation of the experimental points.

Transcriptional rearrangement during the diauxic growth on glutamate

The transcriptomes from cells growing on media supplemented with either ammonium or glutamate have been analysed in triplicate on Affymetrix microarrays (Yeast 2.0) at T1, T2 and T3 sampling points. 771 differentially expressed genes (DEGs, cut-off log2 FC=1) were found in at least one sampling point, indicating that the expression of a relevant fraction of the yeast genome is affected by the nitrogen source during the diauxic growth.

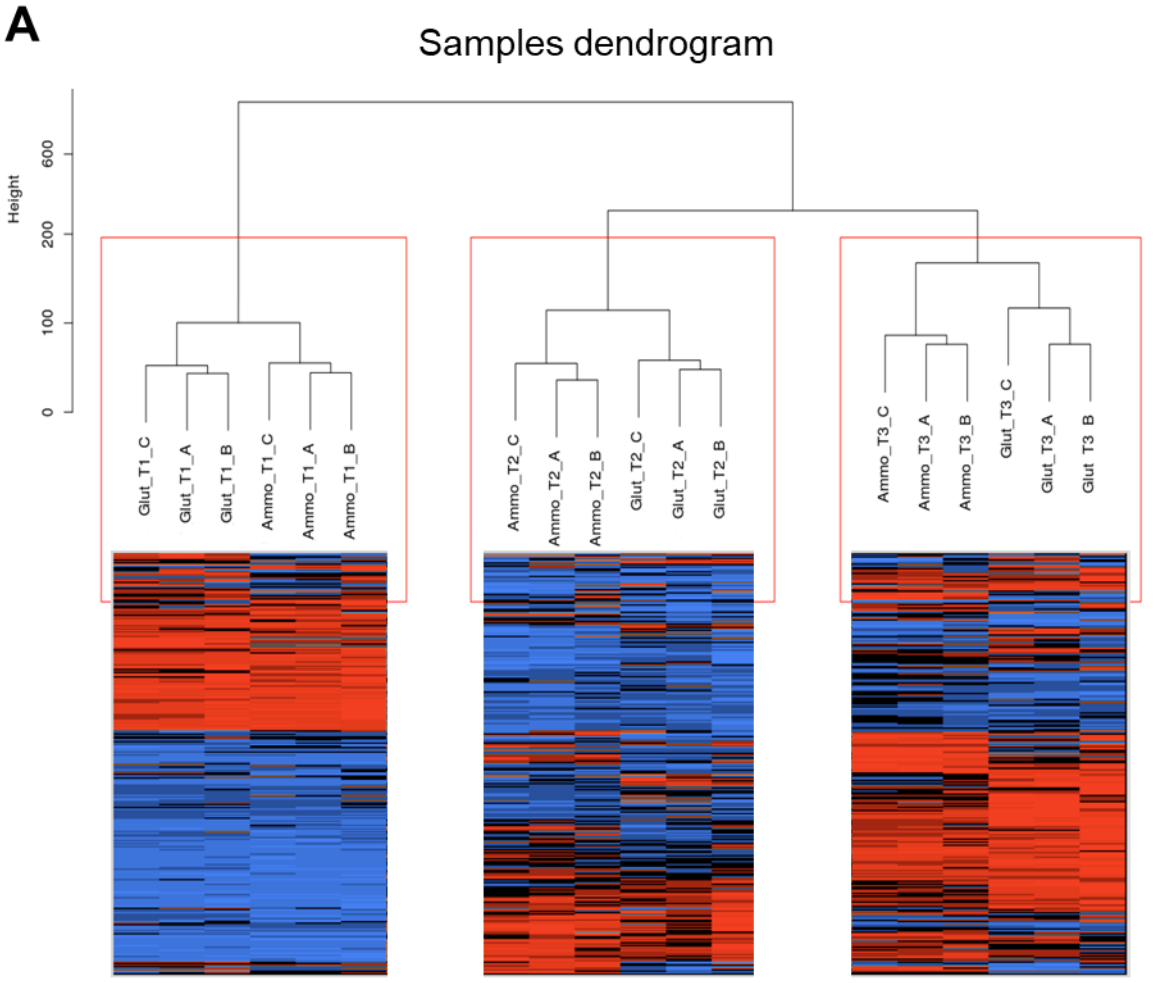
Figure 2A shows that a significant variation in gene expression correlates with the growth phase, the first branching between gene expression data being relative to the different sampling points (Figure 2A). The deepest glutamate-induced rearrangements occur after glucose exhaustion (T2 and T3), suggesting that glutamate affects the transcriptional program in coordination with the physiological state of the cells, highlighting previously undetermined changes between glutamate and ammonium-grown cells (Godard *et al.*, 2007).

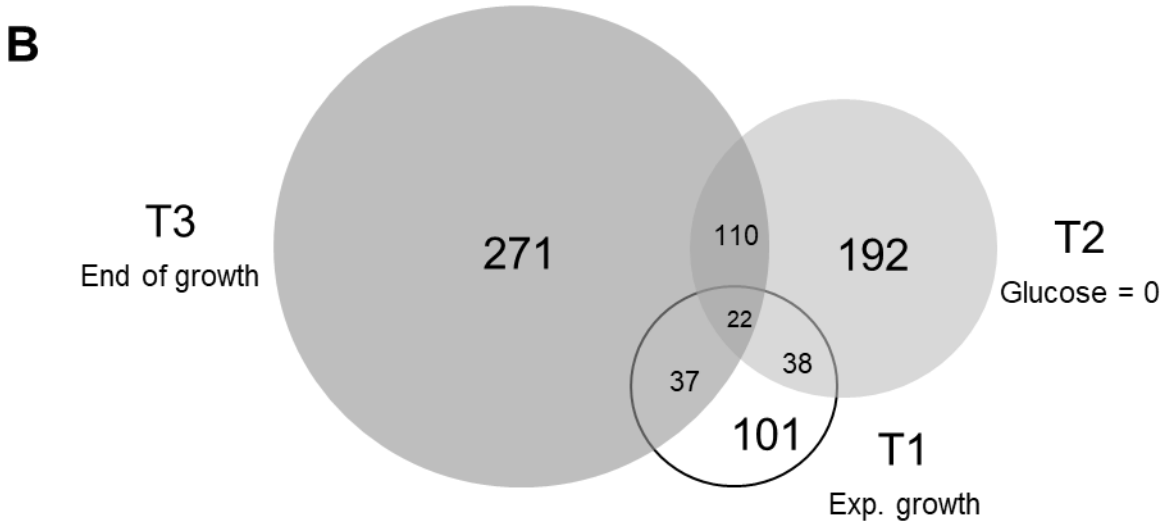
Given the low number of DEGs common among different time points (Figure 2B), we separately searched for enriched KEGG pathways and GO terms connected to the DEGs at each time point (Figure 2C and Supplementary figure S6). For cells exponentially growing on glutamate (T1), we found a overrepresentation of up-regulated GO terms mostly related to amino acid metabolism, consistent with a general de-repression of genes subjected to NCR (nitrogen catabolite repression) (Ljungdahl and Daignan-Fornier, 2012). On the contrary, both GO and KEGG pathways enrichment showed a negative transcriptional response of genes coding for enzymes of the TCA cycle. After glucose exhaustion (T2) the transcriptional reprogramming on glutamate showed a sharp predominance of up-regulated genes. GO and KEGG analyses highlighted the de-repression of gluconeogenic and TCA cycle genes, together with genes related to reserve carbohydrate and fatty acid metabolism and a down-regulation of GO terms related to hexose transporters. Finally, up-regulation of peroxisomal and lipid metabolism was maintained also for T3.

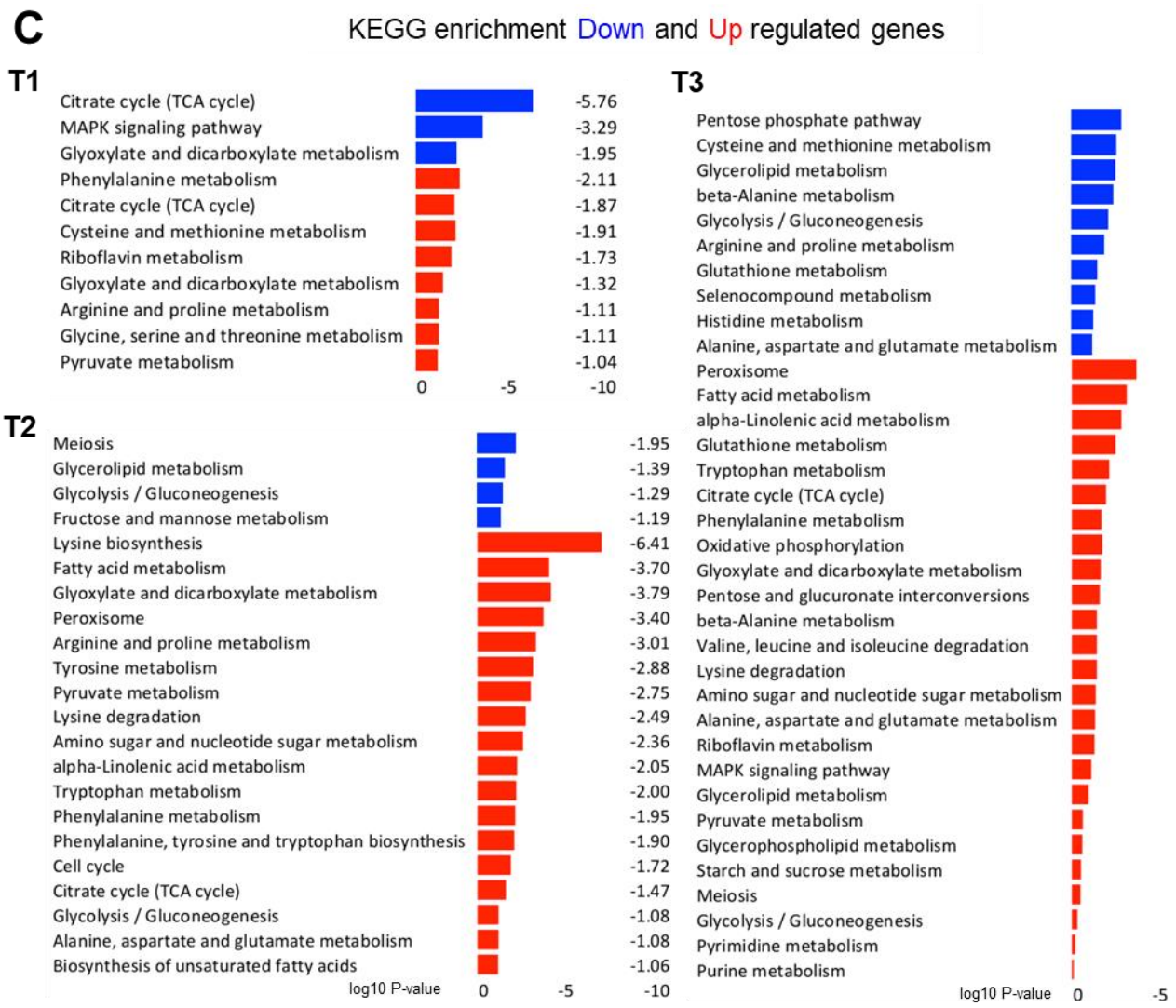
Only 22 DEGs are common to the three sampling points (Figure 2B). These common DEGs could possibly represent the core of the genes involved in the response to glutamate growth. Classification according to GO Biological Functions (Supplementary figure S1) indicates that 9 of the core genes encode proteins implicated in the transport of various metabolites, including amino acids (*TAT2*, *MMP1*), GABA (*UGA4*) and iron (*FET3*). Other functions represented include lipid metabolism (*INO1*, *YSR3* and *ICL2*), aromatic amino acid metabolism (*ARO9*, *ARO10*), and energy sensing (*TOS3*, *SPL2*). Four of the core DEGs have still unknown functions (YBR056W, YKL107W, YDL241W, YGL262W).

Given the observed large number of glutamate responsive genes, we decided to further test the effective relation among the 22 core DEGs. Further analysis, conducted with the STRING software, showed some statistically relevant interactions (co-expression, physical interaction, genomic co-localization) among some of the products of these 22 genes and central cellular functions like gluconeogenesis and energy sensing (Figure 2D). Using the tool "Rank by TF" within the web tool Yeastract, we obtained a ranking of the main transcription factors (TFs) responsible for the regulation of the 22 core glutamate DEGs. Figure 2D indicates that the transcriptional regulatory networks emerging from this analysis are densely interconnected, further sustaining the notion that a common pattern is embedded in the 22 common DEGs.

GO-term enrichment of the 771 genes through the web tool REVIGO provided 2 different tree maps for biological process (Supplementary figure S3) and molecular functions (Supplementary figure S4). Enriched biological processes terms include ketone metabolism, organic substance transport, iron homeostasis, and external encapsulating structure organization. Enriched molecular functions are mainly connected with transporters and oxidoreductase activity, with the involvement of redox cofactors.


A
Samples dendrogram
Height
0
100
200
600
Glut_T1_C
Glut_T1_A
Glut_T1_B
Ammo_T1_C
Ammo_T1_A
Ammo_T1_B
Ammo_T2_C
Ammo_T2_A
Ammo_T2_B
Glut_T2_C
Glut_T2_A
Glut_T2_B
Ammo_T3_C
Ammo_T3_A
Ammo_T3_B
Glut_T3_C
Glut_T3_A
Glut T3 B


B
T3
End of growth
271
110
192
T2
Glucose = 0
22
37
38
101
T1
Exp. growth


C
KEGG enrichment Down and Up regulated genes
T1
Citrate cycle (TCA cycle) -5.76
MAPK signaling pathway -3.29
Glyoxylate and dicarboxylate metabolism -1.95
Phenylalanine metabolism -2.11
Citrate cycle (TCA cycle) -1.87
Cysteine and methionine metabolism -1.91
Riboflavin metabolism -1.73
Glyoxylate and dicarboxylate metabolism -1.32
Arginine and proline metabolism -1.11
Glycine, serine and threonine metabolism -1.11
Pyruvate metabolism -1.04
0 -5 -10
T2
Meiosis -1.95
Glycerolipid metabolism -1.39
Glycolysis / Gluconeogenesis -1.29
Fructose and mannose metabolism -1.19
Lysine biosynthesis -6.41
Fatty acid metabolism -3.70
Glyoxylate and dicarboxylate metabolism -3.79
Peroxisome -3.40
Arginine and proline metabolism -3.01
Tyrosine metabolism -2.88
Pyruvate metabolism -2.75
Lysine degradation -2.49
Amino sugar and nucleotide sugar metabolism -2.36
alpha-Linolenic acid metabolism -2.05
Tryptophan metabolism -2.00
Phenylalanine metabolism -1.95
Phenylalanine, tyrosine and tryptophan biosynthesis -1.90
Cell cycle -1.72
Citrate cycle (TCA cycle) -1.47
Glycolysis / Gluconeogenesis -1.08
Alanine, aspartate and glutamate metabolism -1.08
Biosynthesis of unsaturated fatty acids -1.06
log10 P-value 0 -5 -10
T3
Pentose phosphate pathway
Cysteine and methionine metabolism
Glycerolipid metabolism
beta-Alanine metabolism
Glycolysis / Gluconeogenesis
Arginine and proline metabolism
Glutathione metabolism
Selenocompound metabolism
Histidine metabolism
Alanine, aspartate and glutamate metabolism
Peroxisome
Fatty acid metabolism
alpha-Linolenic acid metabolism
Glutathione metabolism
Tryptophan metabolism
Citrate cycle (TCA cycle)
Phenylalanine metabolism
Oxidative phosphorylation
Glyoxylate and dicarboxylate metabolism
Pentose and glucuronate interconversions
beta-Alanine metabolism
Valine, leucine and isoleucine degradation
Lysine degradation
Amino sugar and nucleotide sugar metabolism
Alanine, aspartate and glutamate metabolism
Riboflavin metabolism
MAPK signaling pathway
Glycerolipid metabolism
Pyruvate metabolism
Glycerophospholipid metabolism
Starch and sucrose metabolism
Meiosis
Glycolysis / Gluconeogenesis
Pyrimidine metabolism
Purine metabolism
log10 P-value 0 -5

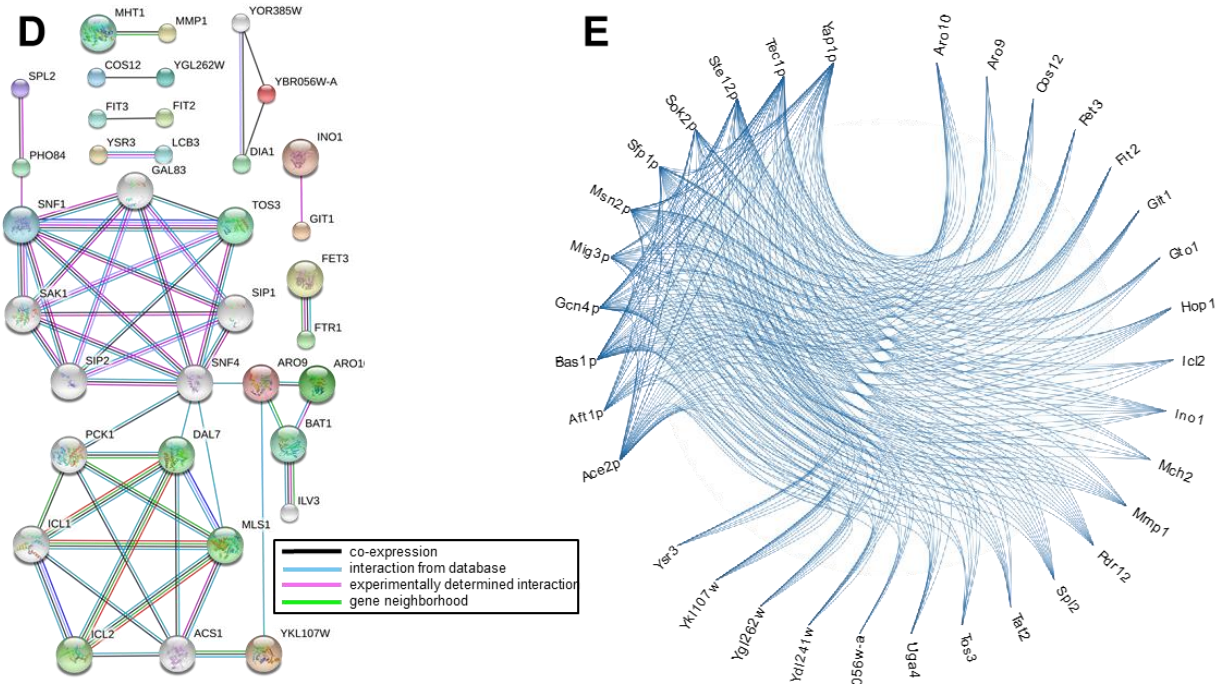


Figure 2 – (A) Correlation treemap of samples similarities. (B) Venn diagram showing the overlay of the differential expressed genes (DEGs) between cells growing on glutamate vs. cells growing on ammonium sulphate in the 3 growth time points. Overall, 771 DEGs are identified. Only 22 DEG are common to the 3 conditions, while 110 DEGs are shared between T2 and T3. 771 DEG, distributed in 3 physiological conditions.
(C) KEGG enrichments of metabolic pathways up or downregulated by the nitrogen source. KEGG Pathway enrichment of up- and down-regulated genes between cells growing on glutamate vs. cells growing on ammonium sulphate in the 3 growth time points. (D) Putative interactions between the 22 core DEGs. Results, obtained with the software STRING showed some cluster of co-expression, gene interaction and genetic neighbourhood. (E) Network of regulators controlling the expression of the 22 core DEGs.

Rewiring of metabolism for post-diauxic growing cells stimulates fatty acid metabolism and the formation of trehalose from glutamate through gluconeogenesis

The transcriptional analysis reported above indicates that growth on glutamate causes a complex, growth phase-modulated transcriptional reprogramming. First of all, we can observe changes in the expression of genes encoding enzymes involved in nitrogen metabolism or in glutamate and related amino acids biosynthesis. Moreover, the transcriptional rewiring includes genes whose products are involved in nutrient transport, stress response, lipid metabolism, oxidoreductase activity, sulphur metabolism response, as well as gluconeogenesis and TCA.

To further characterize metabolic rearrangement induced by growth in glutamate-supplemented medium, intracellular extracts from samples collected at times T1, T2, T3 and T_{Etoh} were analysed by GC-MS (gas chromatography coupled to mass spectrometry) (Figure 3). Parallel samples were analysed for expression of relevant genes, in order to confirm Affymetrix data and to include the T_{Etoh} sampling point. Figure 4 summarizes some of the pathways modulated by glutamate, and that could play a role in the observed enhanced growth phenotype. Blue metabolites in the figure indicate the so-called reporter metabolites, i.e. those spots in metabolism where a substantial transcriptional regulation occurs. That is either to maintain homeostasis or to adjust the concentration of the metabolite to a level required for proper functioning of the metabolic network (Patil and Nielsen, 2005). The full list of reporter metabolite is reported in Supplementary figure S7 and computational details in Materials and methods. Finally, some relevant enzyme activities were measured.

Glutamate-fed cells have high levels of intracellular glutamate, as well as a high content, especially at the T2 and T_{Etoh} time points, of TCA/glyoxylate intermediates (citrate, malate, glyoxylate). An increased content of free fatty acids was detectable at T_{Etoh} in the presence of glutamate, as well as an accumulation of trehalose toward the end of growth.

Several among the identified reporter metabolites take part in the central carbon metabolism, for example TCA/glyoxylate cycle intermediates, and in the biosynthesis of storage macromolecules such as fatty acids and trehalose.

Figure 4 also reports the heatmaps of some enzymatic activities relevant for the pathways identified with previously described techniques. To maintain figure consistency, we expressed data as fold change of the chosen activity on glutamate against ammonium, sampled at the usual times, while respective activities are reported in the Supplementary figure S8. Although sometimes modulated slightly differently

compared to transcriptomic data, all the enzyme assayed resulted in good accordance with the global picture.

As a result of the described multi-layer analysis, we could therefore identify a core metabolic pathway that suggests a possible route for utilization of the glutamate carbon backbone. Briefly, during exponential growth on glucose, the presence of glutamate causes repression of the genes coding for the first three enzymes of the TCA cycle (CIT1, ACO1, IDH1/2), suggesting that in this growth condition anaplerosis of AKG derived by glutamate catabolism could lead to a decreased fuelling of the TCA cycle with glycolytic carbon. In turn, the reduction of acetyl-CoA request by glyoxylate cycle would allow this intermediate to be redistributed along other pathways, like FA metabolism. Following glucose exhaustion, the overexpression of the genes coding for cytosolic malate dehydrogenase (*MDH3*), phosphoenolpyruvate carboxykinase (*PCK1*) fructose-1,6-bisphosphatase (*FBP1*) and of the genes constituting the machinery for biosynthesis of fatty acids (*ACC1, FAS1, FAS2*) suggests an enhanced gluconeogenic flux funnelling the biosynthesis of storage macromolecules. Overexpression of peroxisomal genes for the β-oxidation of fatty acids starts as well after glucose exhaustion and persists during growth on ethanol, in coordination with the expression of genes coding for trehalose and glycogen biosynthesis and catabolism. The apparently counterintuitive activation of both production and catabolism of fatty acid could be useful to the cell for recycling of NADH and NADPH produced in the TCA cycle.

This mechanism has been further validated by the determination of fatty acids and trehalose levels along the growth. Indeed, a cytofluorimetric analysis of the average fluorescence of cells stained with fluorescent marker Nile red, showed an accumulation trend for cells growing in presence of glutamate (Figure 5A). Microscopic analysis revealed a gradual accumulation of lipid droplets, starting from T2, that grew in number and in dimension till the end of growth (Figure 5B). More generally, lipid metabolism and increase of biomass on glutamate seems to be correlated, since the phenotype resulted abolished in the presence of sub-lethal amounts of the inhibitor of fatty acid biosynthesis cerulenin (Figure 5D).

Presence of glutamate stimulates also reserve carbohydrate accumulation. In particular, we evaluated the intracellular accumulation of trehalose. Consistently with literature data, trehalose accumulated during the post-diauxic shift. However, we observed a different trend for the two nitrogen sources, with cells on glutamate showing a greater accumulation, followed by partial reuse, in the time window coincident to the largest difference in biomass formation (Figure 5C). Interestingly, a NMR analysis cells growing on ^{13}C labelled glutamate revealed that a substantial amount of the carbon in trehalose derive from glutamate (Supplementary table S1).

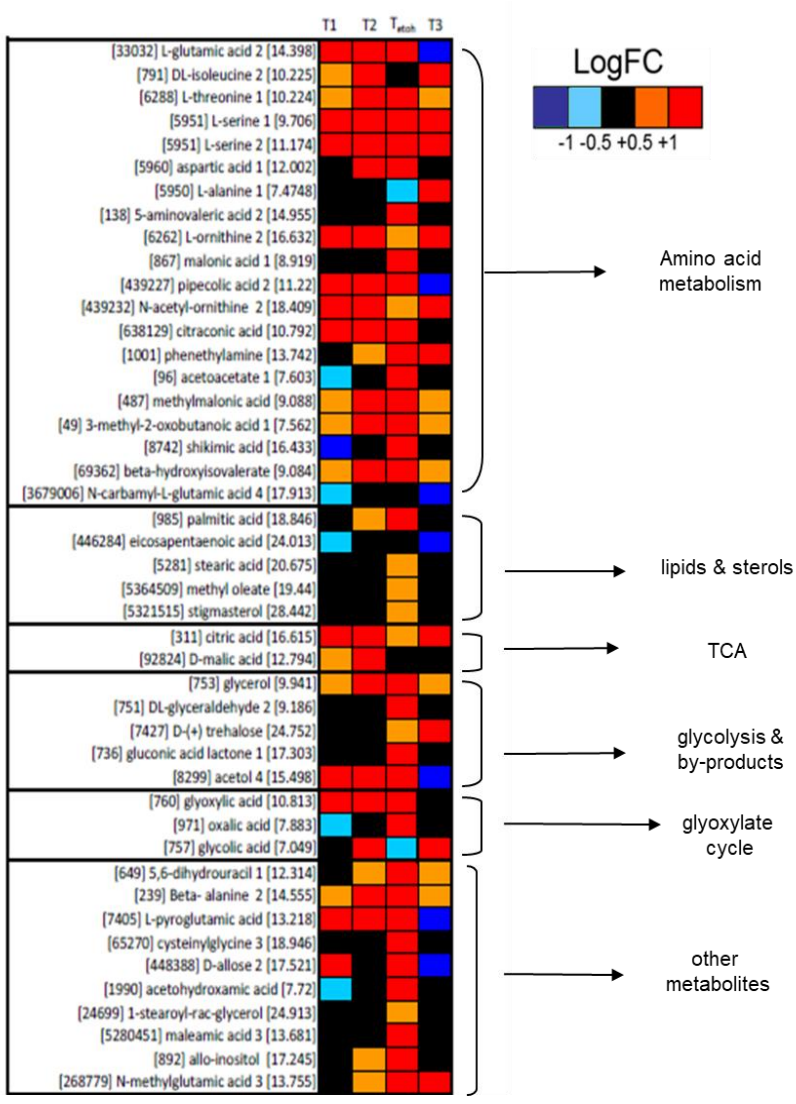


Figure 3 - Heatmap of all the metabolites analysed with the GC-MS, colour coded according to the fold change in glutamate vs ammonium in the different time points and grouped according to their involvement in metabolic pathways.

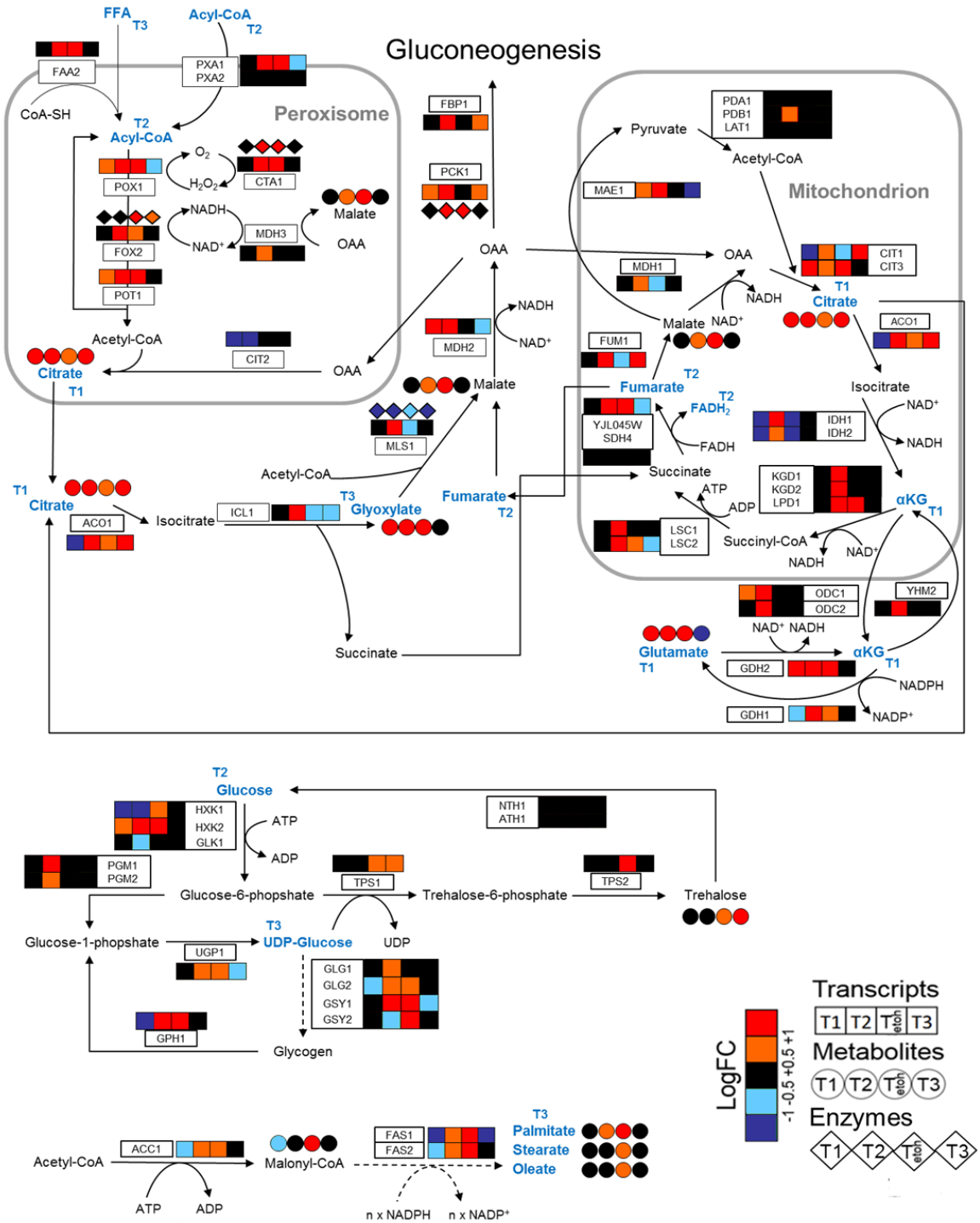


Figure 4 - Landscape of effects induced by glutamate presence. Pathways involved in glutamate assimilation, as resulted by the combination of transcriptomic (squares), metabolomic (circles) and reporter metabolites (text) analysis. For transcriptomic and metabolomic data, the colour code related to Fold Change magnitude is indicated.

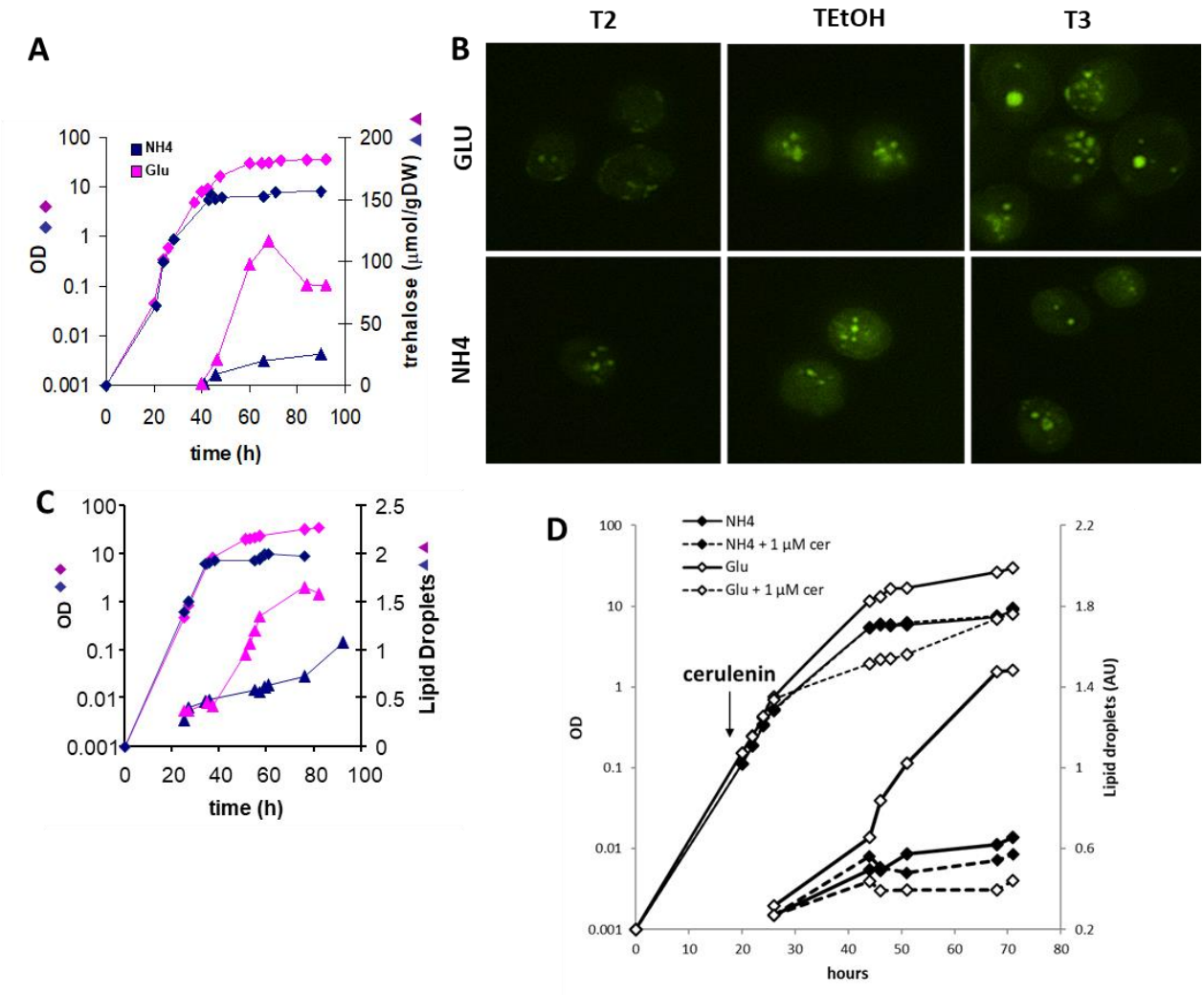


Figure 5 – (A) Accumulation of trehalose in cells with glutamate as nitrogen source approaching the stationary phase. Nitrogen source has a deep impact on trehalose content. Cell extract obtained during growth on ammonium sulphate (black symbols) or glutamate (white symbols) have been subjected to NMR analysis. Trehalose accumulates starting from glucose exhaustion with different dynamics in the two conditions.
(B) Fluorescence microscopy analysis highlights lipid droplets accumulation in the presence of glutamate. Sample cells on ammonium sulphate or glutamate have been collected at the end of growth (T3) and stained with Nile Red (10 µg/ml) for the detection of lipid droplets. Consistently with cytofluorimetric assay, cells on glutamate shows higher amounts of lipid droplets.
(C) Lipid enrichment in the presence of glutamate. Fluorescence induced by Nile Red has been measured during batch grown on glucose thoughtfully the growth curve. Average fluorescence has been weighted for cellular dimension and plotted against time. Presence of glutamate (white symbols) result in an increase of fluorescence starting from glucose exhaustion (40-45 hrs) respect to control cells (black symbols).
(D) Cerulenin inhibition of biomass formation on glutamate.

Central carbon metabolism accounts for the differences in growth yields, which mainly relies on FADH oxidation and glyoxylate

To further confirm and better detail the indications about the metabolic rewiring taking place in the presence of glutamate provided by transcriptome and metabolome data analysis, we performed constraint-based simulations of yeast metabolism for different nitrogen sources.

Constraint-based models rely on a steady state assumption for internal metabolites and are therefore well suited to mimic chemostat experiments.

First, we assessed whether a core model of central carbon metabolism was able to correctly predict the growth yield (dry weight production over glucose consumption rate) given constraints on exchange fluxes with the environment, as derived from chemostat experiment. For this purpose, we exploited our previously developed YENIM model (Frascotti et al, in preparation), which includes the main pathways leading to biomass formation from glucose, glutamate and ammonium.

Different constraints were set for simulations of growth with glutamate or ammonium as nitrogen source, according to corresponding chemostat experiments at different dilution rates D. For glucose, glutamate and oxygen uptake, as well as for ethanol secretion, the exact experimental flux value is imposed. For AKG secretion, the experimental flux is imposed for simulations on glutamate only. For every imposed flux the computational value (lines) and the experimental value (diamonds) always coincide. The availability of ammonium (NH_3 when YENIM1 is used) is bounded according to its concentration in the culture medium times the medium inflow rate. The effective usage (lines) of NH_3 according to FBA (maximization of growth) of YENIM1 model against allowed availability (diamonds) is reported (Figure 6, top right panel). This confirms that NH_3 is not limiting in simulations, in accordance with experimental evidence.

We performed Flux Balance Analysis to predict the maximal growth yield of the two experimental conditions, given experimental constraints on the influx of glucose, glutamate and ammonium and on the efflux of main secreted products: ethanol and AKG.

Figure 7A shows that the core model predicts experimental growth yields as accurately as the genome-wide model YEAST 7 (Aung, Henry and Walker, 2013). Accuracy of both models worsen at high dilution rates.

During a chemostat culture, the variation of the dilution rate (i.e. of the growth rate) shifts the metabolism from completely respiratory to a mixed respiro-fermentative condition, characterized by a progressive glucose accumulation and ethanol production.

It can be noticed (Figure 6) that, the growth yield on glutamate largely outperforms that on ammonium, especially when metabolism is completely respiratory, and that ethanol production occurs at lower dilution rates in glutamate. Conversely, the spread between the growth yields of the two conditions decreases as the dilution rate increases, with the growth yield on glutamate, as well as the accumulation of ethanol and AKG, reaching levels comparable to those reported for shaking flask data. In light of these results, we are confident that central carbon metabolism is strongly involved in the rearrangement and that we can focus on the simulation outcomes of the more handily core model for more in-depth investigations in the following.

To assess which pathways are mainly exploited in the different nitrogen conditions, we performed single reaction deletion analysis. Reactions that, when depleted, mostly affect the capability to either abide by experimental constraints or to reach experimental growth yields indicate the pathways the metabolism mainly rely on.

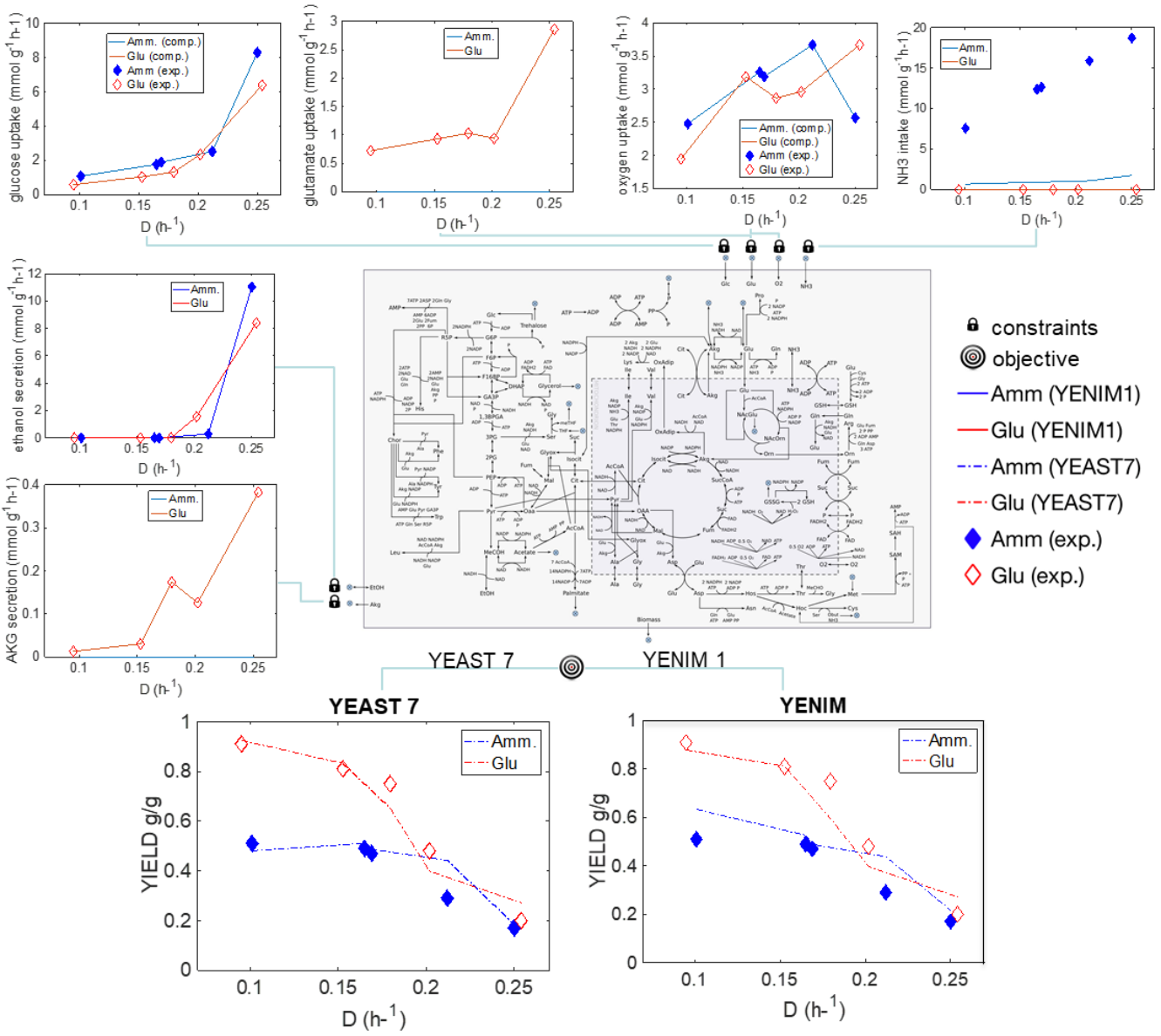


Figure 6 - Core model versus genome-wide model growth yield predictions. Constraint-based simulation of chemostat experiments for different growth conditions: i) ammonium and glucose (aerobic), ii) glutamate and glucose (aerobic) iii) ammonium and glucose (anaerobic). Optimal growth yield (biomass over glucose) as predicted by FBA - for both YEAST 7 genome-wide model and YEAST 1 core model – given constraints on nutrient consumption and secretion, as compared to experimental growth yield.

Figure 7 report those reactions that display differential effect for growth on ammonium sulphate and glutamate at dilution rate 0.2 (to simulate a fermentative condition). Obvious differences directly follow from the different nitrogen source: glutamate uptake is essential in glutamate (EX_Glu), whereas ammonium uptake is essential in ammonium (EX_NH3). Along similar lines, NADPH-dependent glutamate dehydrogenase is essential in ammonium as it is needed to synthesize glutamate for biomass. Similar results are retrieved from analysis at different dilution rates (Supplementary figure S10).

A large difference emerged in the utilization of succinate dehydrogenase (Sdh) and $FADH_2$ oxidation in the respiratory chain (Resp2). While cells in ammonium are completely insensitive to removal of this pathways, cells in glutamate largely depend on them. Notably, the extent of oxygen utilization (EX_O2) does not have a significantly different impact on the two conditions.

The cytosolic reaction that converts isocitrate to succinate and glyoxylate (Icl) seems to play a significant role in glutamate, whereas it is negligible in ammonium.

Also, the reaction from pyruvate to acetaldehyde (Pdc) affects growth on glutamate at a larger extent. Worth of note, the glycolytic reaction catalysed by aldolase (Fba) is only essential in glutamate.

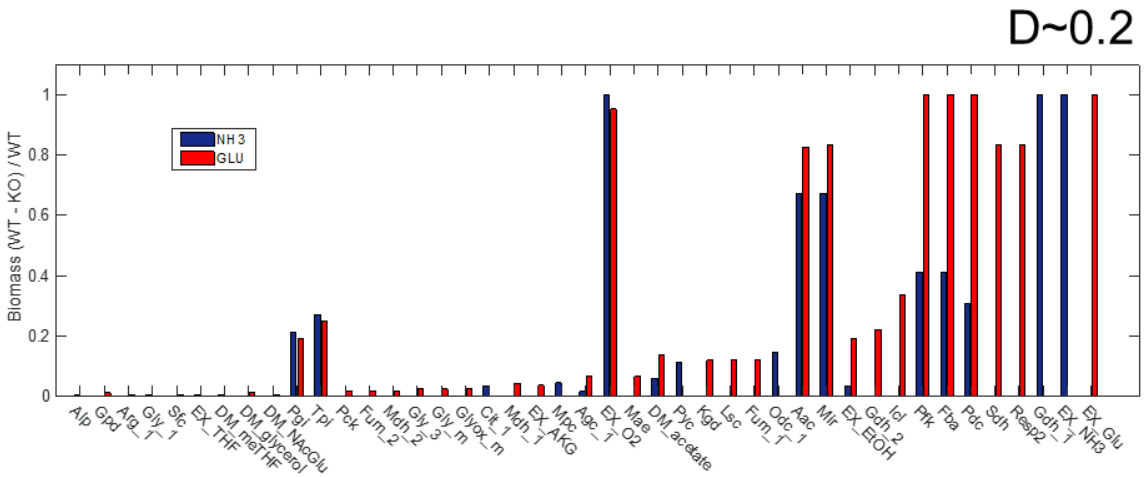


Figure 7 - Reaction deletion analysis (D~0.2): the fraction of biomass reduction observed after deleting a single reaction, while retaining the wild-type constraints, is reported for reaction having a different effect in glutamate and ammonium (aerobic) simulated growth conditions

Interplay between glutamate and ethanol assimilation improves growth yields

Batch experiments showed a more complex behaviour as compared to chemostat experiments, with utilization of glutamate becoming more relevant when glucose is no longer available.

We hypothesize that major differences between the two experiments are associated with the different extent of utilization of the respiratory chain. It is indeed well known that high glucose levels repress mitochondrial activity (Busti *et al.*, 2010) (Kayikci and Nielsen, 2015). On the other hand, during glucose-limited chemostat growth, the residual glucose does not exceed the concentration needed to establish glucose repression. The onset of respiro-fermentative metabolism at high dilution rates is the result of the saturation of the oxidative pathways (Käppeli, 1981) and of the uncoupling of respiration due to organic acid production (Postma *et al.*, 1989).

We analysed *in silico* the impact of mitochondrial capacity on the ratio between the maximal growth yield on glutamate over that on ammonium, when the carbon source is either ethanol or glucose. At this aim, we varied the upper bound of the oxygen consumption rate, which is a direct proxy of the respiratory chain capacity, from 1 to 5 mmol/h, given 1 mmol/h of either glucose or ethanol consumption.

A first notable result comes from the observation that the spread between the theoretical maximal growth yield of the two nitrogen conditions is not significantly affected by mitochondrial capacity, when glucose is used as carbon source (Figure 8A). This result suggests that the reason for a larger difference between the two nitrogen sources - in terms of growth yields - for low dilution rates does not merely lay in the enhanced utilization of the respiratory chain, but rather in a multi-faced interaction of factors.

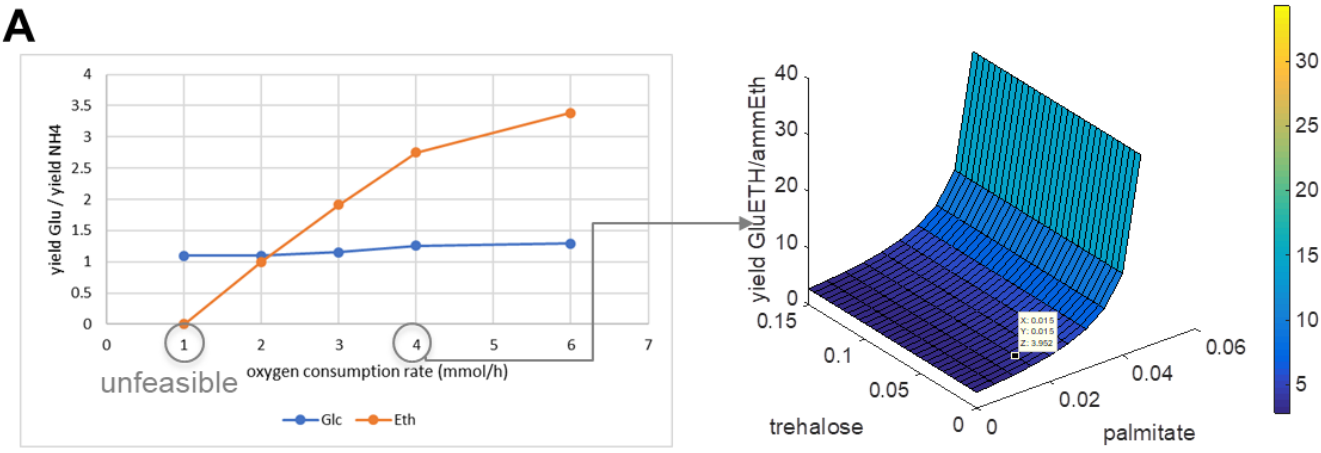
In support of this hypothesis, we observed that, when releasing a single constraint at the time, the remaining constraints suffice for a correct prediction of the growth yields. For example, when the constraint on AKG secretion was released, we still observed secretion of AKG in the presence of glutamate (Supplementary figure S9A). Along similar lines, when the constraint on glucose uptake was released, we still observed a good correspondence between computational and experimental glucose uptake rates (Supplementary figure S9-B). Even when both constraints on carbon secretion (in the form of AKG or ethanol) were released, the accuracy of the growth yield predictions was not significantly affected (Supplementary figure S9-C). Notably, when the constraint on glucose uptake and AKG secretion are concurrently released, the spread between the growth yield of the two nitrogen conditions shrinks (Supplementary figure S9-B/D). In this situation, cells on glutamate secrete more AKG and intake more glucose than expected. As the glucose uptake rates of the two conditions get closer, so do the corresponding yields, meaning that advantage on glutamate derived from the capacity to make comparable biomass amounts by using less glucose, and thus by using

glutamate as carbon source. When cells are allowed to secrete more AKG, they will use glucose rather than glutamate as carbon source. This result suggests that utilization of glutamate nitrogen is coupled either with AKG secretion or with lower glucose uptake rates. All in all, these findings clarify that that the small difference in the growth yield, observed in shake-flask, can be ascribed to glucose repression of the utilization of glutamate carbons, which results in high secretion rates of AKG. Taken together these findings indicate a high degree of consistency between batch experiments and computational results for high glucose conditions.

As opposed to utilization of glucose, utilization of ethanol strictly depends on mitochondrial activity: accordingly, when only one mol of oxygen is allowed to be consumed per mol of ethanol no feasible solution could be found (Figure 8A). At least 2 moles of oxygen are required to fully consume 1 of ethanol. A less expected result comes from the observation that the spread between the theoretical maximal growth yield (biomass over ethanol consumption) of the two nitrogen conditions, is instead significantly affected by mitochondrial capacity, when ethanol (rather than glucose) is used as carbon source.

For instance, when 4 moles of oxygen can be consumed per mol of ethanol, the growth yield on glutamate outperforms by about 3 times (2.8) that on ammonium, indicating that some oxygen is employed to metabolize glutamate.

Indeed, growth of ammonium and glutamate yeast cultures in the absence of oxygenation and with the presence of the respiratory chain inhibitor Antimycin A completely abolishes enhanced growth (Supplementary figure S12).



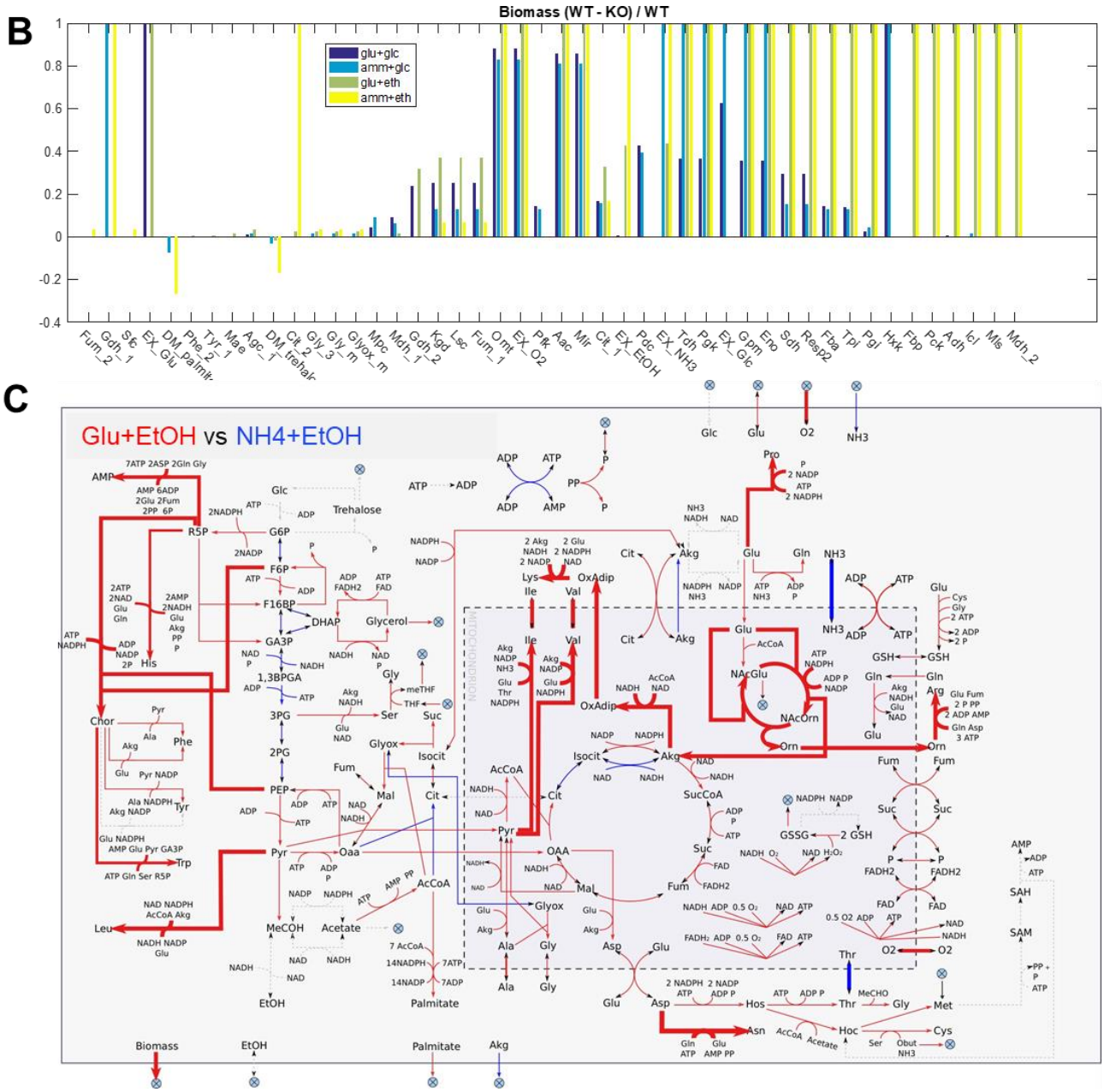


Figure 8 – Interplay of carbon, nitrogen and oxygenation is crucial for the enhanced growth phenotype.
(A) Simulation of growth on glutamate and either glucose or ethanol as carbon source. FBA was performed with different upper bounds of oxygen consumption (left). The value of 4 mmol/h was used to simulate growth on glutamate and ammonium with ethanol and with forced production of palmitate and trehalose with different lower bounds (right)
(B) Reaction deletion analysis for growth with forced palmitate and trehalose production
(C) Analysis of differential fluxes between growth in glutamate+ethanol and ammonium+ethanol with forced production of palmitate with the limit of at least 80% of the optimal growth yield.

Glutamate-induced metabolic rewiring does not only benefit biomass formation, but also stress resistance

As long as optimal solutions are concerned we did not observe differential production of fatty acids nor trehalose between glucose and ammonium, as opposed to what observed in previous experimental analysis. This inconsistency suggests that secretion of these metabolites is not optimal for biomass and that the system may be in a suboptimal space.

To investigate this issue, we compared the two nitrogen conditions, when some secretion of both palmitate and trehalose is forced. We focused on a condition in which glutamate largely outperforms ammonium (namely the point corresponding to oxygen 4 in Figure 8A – left), and we scanned different values of the lower bounds of the production of the two metabolites. Figure 8A (right) reports the ratio between the growth yields of the two conditions for each pair of values. Remarkably, the spread between the two yields tend to increase as palmitate and trehalose accumulation increases. For instance, when we forced a 0.015 accumulation of both trehalose and palmitate the yield ratio reached 4. We focused on this latter condition to analyse the different metabolic routes followed by the cells, when either ethanol or glucose are supplied in the presence of either glutamate or ammonium. At this aim we compared the effects of a single reaction deletion analysis.

Figure 8B reports the effect on growth for those reaction deletions whose effect differ in at least one of the 4 conditions under study. It can be noticed that the deletions of palmitate and trehalose restore the optimal state and thus result in a growth yield increase, which is more significant in ammonium than in glutamate. Interestingly, the gly_3 reaction (Glyox[m] + Ala[m] => Gly[m] + Pyr[m]) is essential in ethanol only when ammonium is the nitrogen source, whereas Cit_1 (Oaa[m] + AcCoA[m] => Cit[m]) seems to play a more significant role in glutamate plus ethanol.

We analysed the fluxes that significantly differ between ammonium and glutamate, when ethanol is the carbon source. We sampled the flux distributions that allow to achieve at least 80% of optimal growth (given the constraints on trehalose and palmitate accumulation). We used random objective functions as sampling method, as in (Damiani *et al.*, 2014). This kind of sampling is not limited only to biomass yield and can give more information on our system.

It is apparent (Figure 8C) that in the glutamate condition there is a higher activation of the glyoxylate cycle and of amino acids metabolism, in agreement with transcriptional data analysis results. Although the same lower bound of palmitate and trehalose accumulation was imposed to both nitrogen conditions, we observed a significantly higher accumulation of palmitate in glutamate. Conversely, trehalose production does

not differ in the two conditions. This result was expected given that trehalose production is demanding in terms of ATP consumption but, as opposed to palmitate secretion, is not involved in redox balance and thus is inefficient not only for growth but for every (random) metabolic function.

It can be hypothesized that, when glucose is the carbon source, because utilization of glutamate for biomass does not provide a significant advantage in growth yield, the systems do not rewire completely towards glutamate utilization, but exploits glutamate for storage of fatty acids and trehalose which may in turn provide advantages that are not related to biomass formation, as for instance resistance to stress, and hence are not taken into account by FBA computations.

To investigate the physiological reason for the higher production of storage molecules of cells growing on glutamate, we assayed whether these cells were resistant to various stresses. From the drop-test in Figure 9 we can observe that glutamate-grown cells show a lower sensitivity, compared to ammonium-grown cells to heat, osmotic, saline and oxidative stress. The latter was confirmed by liquid cultures treated with oxygen peroxide (Supplementary figure S11).

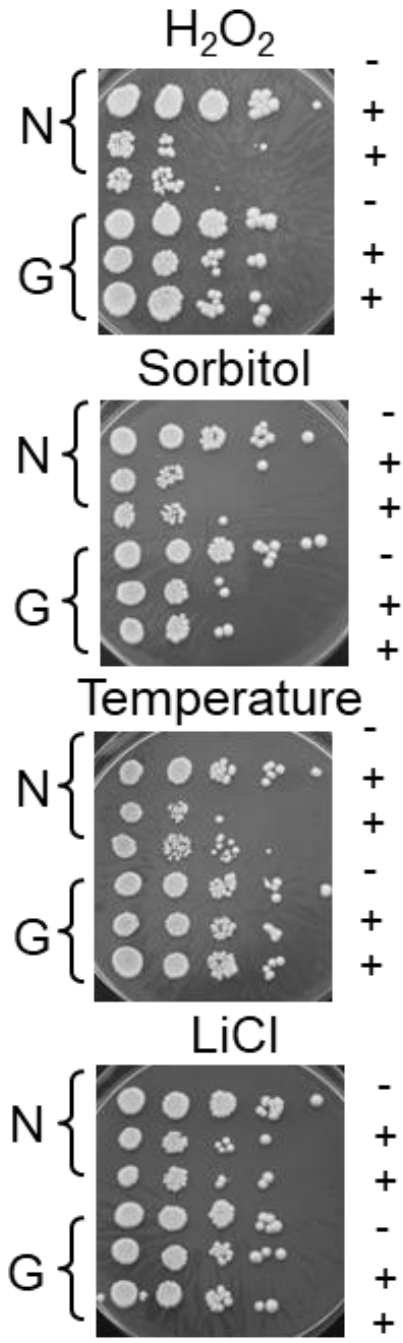


Figure 9 - Glutamate confers stress resistance to exponentially growing cells. Aliquots (0.5 OD) of exponentially growing cells have been collected and exposed to the following stresses: hydrogen peroxide 2 mM (1 hr), 50°C (10 minutes), LiCl 2M (3 hr), Sorbitol 4 M (1 hr). Sequential dilution of treated and untreated cells were spotted on YPD and incubated for 3 days. N: ammonium sulphate, G: glutamate, -: untreated cells, +: treated cells.

Discussion

S. cerevisiae, like every organism, apply different growth strategies to exploit what the environment can offer, and is able to rewire a lot of different, although interconnected cellular processes to react appropriately to changes in the available carbon- and nitrogen-containing compounds, as well as other micronutrients and to perturbation of the cellular energy status (Zaman *et al.*, 2008).

Yeast adaptation to different nitrogen sources relies on many sensing mechanisms that are able to discriminate the quantity and quality of both intracellular and extracellular nitrogen. Cells react to the growth in different nitrogen sources by activating different pathways and modulating transcription and metabolism at different levels (Cooper, 1982) (Ljungdahl and Daignan-Fornier, 2012).

Godard and colleagues investigated the effects of 21 nitrogen sources on yeast growth rate and transcription (Godard *et al.*, 2007). They were able to classify most of the nitrogen sources into two groups, which they called A and B. Compounds in the group A allow fast growth and show activation of the nitrogen catabolite repression (NCR) and of the unfolded protein response (UPR), while they display a mild activation of SPS pathway and of the retrograde signalling (RTG). On the other hand, the NCR is inactive in the nitrogen sources of the B group, while general amino acid control (GAAC) is active. According to this study, both ammonium, the standard nitrogen source for yeast culture media, and glutamate belong to the group A of nitrogen sources and can therefore be considered among those "preferred" by *S. cerevisiae*. They sustain growth with a similar doubling time and they slightly differ in the activation of UPR, which is weaker in ammonium, and in the regulation of RTG, which is completely inactive in glutamate.

We decided to further expand the comparison between these two nitrogen sources, starting from a more in-depth physiological characterization. A very interesting result that we observed was the very interesting relationship between the growth phase and the assimilation of the different nitrogen source. In fact, growth on glutamate always present very interesting features, compared to growth in ammonium. The difference in cell size observed during the exponential growth on glucose, which is not the result of vacuolation but correlates with an increase in the RNA and protein content, reflects an alteration of the length of the cell cycle phases that deserves further analysis.

Interestingly, the difference in phenotype persists during the subsequent phases of the diauxic growth, but the specific phenotype changes. In fact, after glucose is exhausted, and during growth on the ethanol produced during glucose fermentation, glutamate grown-cells are able to use the extra carbon obtained by glutamate to make biomass, divide more and reach a higher cell density. Moreover, chemostat experiments proved that the metabolic state of the cell has a great influence on the enhanced growth

phenotype of glutamate cells, with a higher difference in biomass yield obtained with growth at low dilution rate (respiratory metabolism) that flatten gradually moving to a higher dilution rate (respiro-fermentative metabolism).

The integrated analysis we performed was crucial to put together the pieces of the puzzle and to understand from different angles the deep metabolic changes that allow cells to perform the apparent easy task that is to exploit the extra carbon backbone of a richer nutrient.

Simulation of chemostat experiments was the best starting point for our computational analysis and allowed us to validate the reliability of our core model in predicting phenotypes related to the presence of different nitrogen sources as well as a genome-wide model. Growth yields at different dilution rates were substantially comparable to experimental values, especially for low dilution rates. Moreover, predictions of relevant reaction carried out with deletion analysis highlighted the importance of respiration. Accordingly, cells grown in anaerobiosis, in the presence of the respiratory inhibitor antimycin A, or with deletion of the succinate dehydrogenase do not display enhanced growth.

One of the possible conclusions derived from the data discussed so far could be that the respiratory metabolism is the only responsible for the observed phenotype. However, during growth in shaking flasks, different factors are at play. Because of the higher concentrations of nutrient, glucose plays a much more important role, influencing not only respiratory enzymes but generally re-shaping cellular transcription and metabolism (Zaman *et al.*, 2009).

Experimental data show an interesting correlation between glutamate-derived AKG secretion and glucose uptake from the medium only during growth on glucose as a carbon source. It seems that cells fail to correctly utilize AKG and, therefore is forced to waste a great part of it. During the exponential growth rate, accordingly, we do not observe any up-regulation of the internal AKG mitochondrial transporters. That begins to appear at glucose exhaustion, together with the up-regulation of TCA cycle. Flux balance analysis further strengthen the link between glucose and AKG, showing that the imposed constrains on the import of the former and on the export of the latter are interconnected and the result of metabolic reactions. In fact, only, abolishing the two of them leads to a loss of accuracy in the predictions of growth yields between ammonium and glutamate. Strikingly, varying oxygen consumption without changing these parameters did not affect growth on glucose.

During simulations on ethanol, on the other hand, yeast metabolism was instead shown to be capable of fully exploiting glutamate carbons without having to excrete them in the form of AKG, and this process was shown to be highly dependent on respiration, confirming experimental data.

Finally, we showed that the accumulation of more biomass is only one of the consequences of the metabolic rearrangement of glutamate-grown cells. Glutamate cultures also enhance the gluconeogenic pathway and redirect fluxes towards the biosynthesis of storage molecules like fatty acids and trehalose, for which we observe an accumulation. The continuous exchange of information between the wet lab analyses and the computational simulations, allowed us to affirm that accumulation of this compounds is not strictly necessary for the enhanced biomass production, from a theoretical metabolic point of view. Nevertheless, simulations with suboptimal biomass yields proved that it is possible to produce trehalose and fatty acids and still maintain a difference between glutamate and ammonium and that. More *in vivo* experiments showed that fatty acid biosynthesis is essential to maintain the enhanced growth phenotype in glutamate, while having little effects on ammonium-grown cells. On the other hand, trehalose derived from glutamate could confer the cells different growth advantages, being implicated in their enhanced stress resistance.

In conclusion, integration of physiology, transcriptomics, metabolomics and computational simulations proved to be a valuable tool for a system-level understanding of yeast reaction to ammonium and glutamate. It could be interesting to further analyse differences between the other nitrogen sources with the same approach both to understand yeast regulation of growth and metabolism and to identify more appropriate growth conditions for cultivation of industrially relevant strains. In this sense, both glutamate-induced yeast enhanced growth and stress resistance could be further evaluated for yeast utilization as a cell factory.

References

Aung, H. W., Henry, S. A. and Walker, L. P. (2013) 'Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism', Industrial Biotechnology, 9(4), pp. 215–228. doi: 10.1089/ind.2013.0013.

Bachmann, H. et al. (2013) 'Availability of public goods shapes the evolution of competing metabolic strategies', Proceedings of the National Academy of Sciences, 110(35), pp. 14302–14307. doi: 10.1073/pnas.1308523110.

Brambilla, L. et al. (1999) 'NADH reoxidation does not control glycolytic flux during exposure of respiring Saccharomyces cerevisiae cultures to glucose excess', FEMS Microbiology Letters, 171(2), pp. 133–140. doi: 10.1016/S0378-1097(98)00589-8.

Broach, J. R. (2012) 'Nutritional control of growth and development in yeast.', Genetics, 192(1), pp. 73–105. doi: 10.1534/genetics.111.135731.

Busti, S. et al. (2010) 'Glucose signaling-mediated coordination of cell growth and cell cycle in Saccharomyces cerevisiae.', Sensors (Basel, Switzerland). Molecular Diversity Preservation International, 10(6), pp. 6195–6240. doi: 10.3390/s100606195.

Cooper, T. G. (1982) Nitrogen Metabolism in Saccharomyces cerevisiae., The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. doi: 10.1101/087969180.11B.39.

Cortassa, S., Aon, J. C. and Aon, M. A. (1995) 'Fluxes of carbon, phosphorylation, and redox intermediates during growth of saccharomyces cerevisiae on different carbon sources', Biotechnology and Bioengineering, 47(2), pp. 193–208. doi: 10.1002/bit.260470211.

Damiani, C. et al. (2014) 'An ensemble evolutionary constraint-based approach to understand the emergence of metabolic phenotypes', Natural Computing. Springer Netherlands, 13(3), pp. 321–331. doi: 10.1007/s11047-014-9439-4.

Damiani, C. et al. (2017) 'A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth: The WarburQ effect', PLoS Computational Biology, 13(9), pp. 1–29. doi: 10.1371/journal.pcbi.1005758.

van Dijken JP et al. (2000) 'An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains', Enzyme and Microbial Technology, 26(9–10), pp. 706–714.

Francois, J. and Parrou, J. L. (2001) 'Reserve carbohydrates metabolism in the yeast Saccharomyces cerevisiae', FEMS Microbiology Reviews, 25, pp. 125–145.

Godard, P. et al. (2007) 'Effect of 21 different nitrogen sources on global gene expression in the yeast Saccharomyces cerevisiae.', Molecular and cellular biology, 27(8), pp. 3065–3086. doi: 10.1128/MCB.01084-06.

Gonzalez, B., Francois, J. and Renaud, M. (1997) 'A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol.', Yeast (Chichester, England). England, 13(14), pp. 1347–1355. doi: 10.1002/(SICI)1097-0061(199711)13:14<1347::AID-YEA176>3.0.CO;2-O.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', Cell. Elsevier Inc., 144(5), pp. 646–74. doi: 10.1016/j.cell.2011.02.013.

de Jong-Gubbels, P. et al. (1995) 'Regulation of carbon metabolism in chemostat cultures of Saccharomyces cerevisiae grown on mixtures of glucose and ethanol', Yeast (Chichester, England), 11(5), pp. 407–418. doi: 10.1002/yea.320110503.

Käppeli, O. (1981) 'Regulation of Carbon Metabolism in Saccharomyces cerevisiae and Related Yeasts', pp. 181–209.

Kayikci, Ö. and Nielsen, J. (2015) 'Glucose repression in Saccharomyces cerevisiae', FEMS Yeast Research, 15(6), pp. 1–8. doi: 10.1093/femsyr/fov068.

Landi, C. et al. (2015) 'High cell density culture with S. cerevisiae CEN.PK113-5D for IL-1β production: Optimization, modeling, and physiological aspects', Bioprocess and Biosystems Engineering, 38(2), pp. 251–261. doi: 10.1007/s00449-014-1264-8.

Ljungdahl, P. O. and Daignan-Fornier, B. (2012) 'Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae', Genetics, 190(3), pp. 885–929. doi: 10.1534/genetics.111.133306.

Miller, S. M. and Magasanik, B. (1990) 'Role of NAD-linked glutamate dehydrogenase in nitrogen metabolism in Saccharomyces cerevisiae', Journal of Bacteriology, 172(9), pp. 4927–4935. doi: 10.1128/jb.172.9.4927-4935.1990.

Parrou, J. L. and François, J. (1997) 'A simplified procedure for a rapid and reliable assay of both glycogen and trehalose in whole yeast cells.', Analytical biochemistry, 248, pp. 186–188. doi: 10.1006/abio.1997.2138.

Patil, K. R. and Nielsen, J. (2005) 'Uncovering transcriptional regulation of metabolism by using metabolic network topology', Proc Natl Acad Sci U S A, 102(8), pp. 2685–2689. doi: 10.1073/pnas.0406811102.

Postma, E. et al. (1989) 'Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae.', Applied and Environmental Microbiology, 55(2), pp. 468–477. doi: 0099-2240/89/020468-10$02.00/0.

Ritchie, M. E. et al. (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', Nucleic Acids Research. Oxford University Press, 43(7), pp. e47–e47. doi: 10.1093/nar/gkv007.

Schmitt, M. E., Brown, T. A. and Trumpower, B. L. (1990) 'A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae', Nucleic Acids Research, 18(10), pp. 3091–3092. doi: 10.1093/nar/18.10.3091.

Smets, B. et al. (2010) Life in the midst of scarcity: Adaptations to nutrient availability in Saccharomyces cerevisiae, Current Genetics. doi: 10.1007/s00294-009-0287-1.

Ueda, M. et al. (1985) 'Peroxisomal Localization of Enzymes Related to Fatty Acid β-Oxidation in an n-Alkane-grown Yeast, Candida tropicalis', Agricultural and Biological Chemistry, 49(6), pp. 1821–1828. doi: 10.1080/00021369.1985.10866984.

Verduyn, C. et al. (1992) 'Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation', Yeast (Chichester, England), 8(7), pp. 501–517. doi: 10.1002/yea.320080703.

Wortel, M. T. et al. (2018) 'Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield', PLoS Computational Biology, 14(2), pp. 1–21. doi: 10.1371/journal.pcbi.1006010.

Zaman, S. et al. (2008) 'How Saccharomyces responds to nutrients.', Annual review of genetics, 42, pp. 27–81. doi: 10.1146/annurev.genet.41.110306.130206.

Zaman, S. et al. (2009) 'Glucose regulates transcription in yeast through a network of signaling pathways.', Molecular systems biology. EMBO Press, 5(1), p. 245. doi: 10.1038/msb.2009.2.

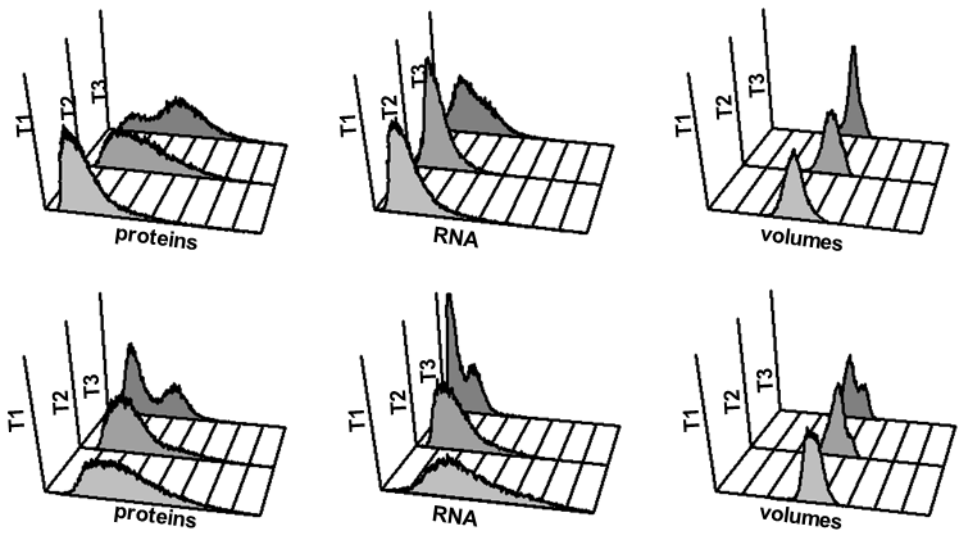
Supplementary information

hours	OD_{660}	Glc (g/l)	Etoh (g/l)	C^{13} Etoh (%)	Trehalose (mmol/OD)	C^{13} Trehalose (%)
40	16.3	0	6.44	< 1	0.0025	0
60	30.3	0	2.32	<1	0.1342	31
84	35.1	0	0	<1	0.1121	29

Supplementary table S1 - NMR determination of trehalose content in cells growing on ^{13}C-glutamate and 2% glucose.

log (h)	OD	Glu (mM)	^{13}C%	AKG (mM)	^{13}C%	Fum (mM)	^{13}C%
0	0.1	85	98	0,13	1	0	0
4	0.3	78	98	0,3	63	0	0
6	0.6	71	98	0,82	85	0	0
24	13.8	46	nd	7,9	95	0.18	25
27.5	16.3	42	nd	8,6	95	0.24	22%
48	30.3	25	91	7,2	95	0.2	22%
53	33.4	27	92	7,2	94	0.2	22%
72	35.1	29	93	7,3	94	0.2	22%

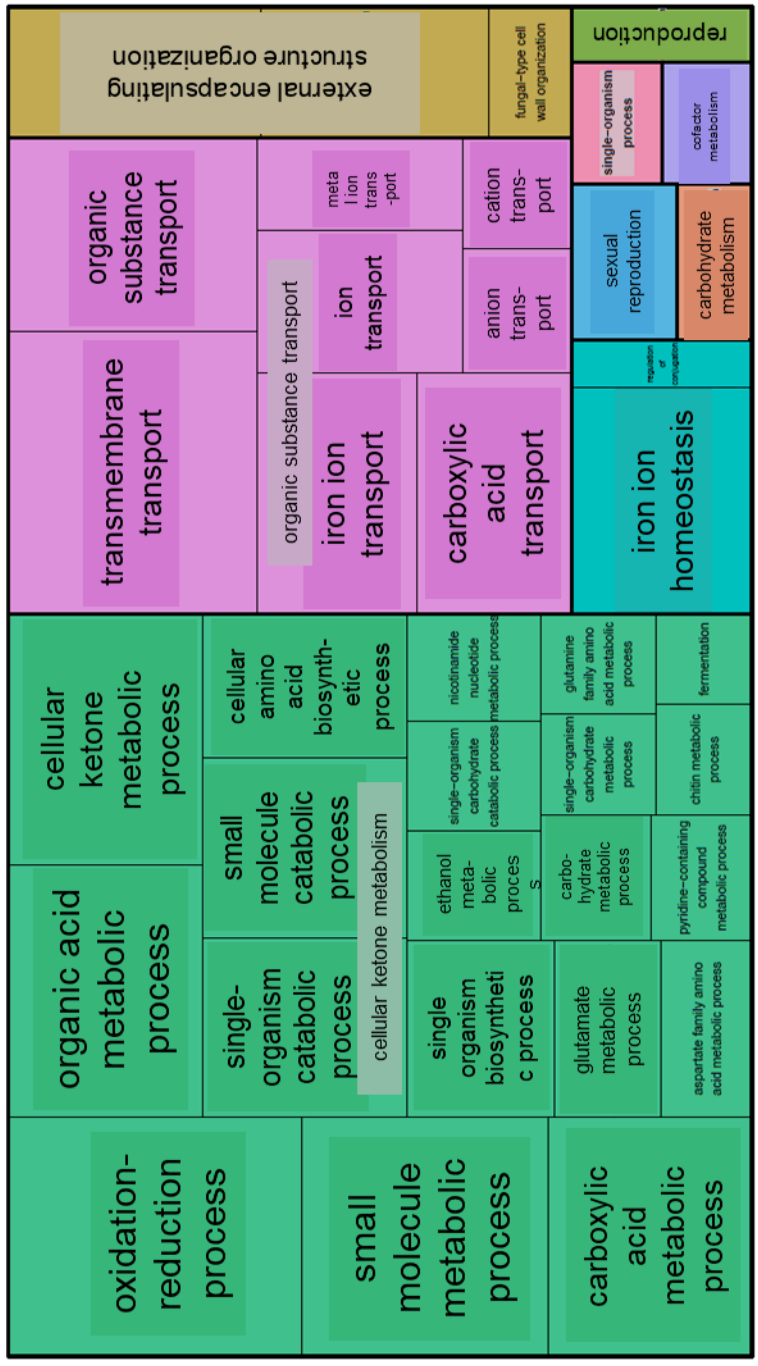
Supplementary table S2 - AKG released in the medium originate from glutamate deamination. Table reports the result of a ^{13}C-glutamate pulse to exponentially growing cells. Few hours after the substitution, AKG resulted labelled in high percentage. Labelling was maintained till the end of growth.



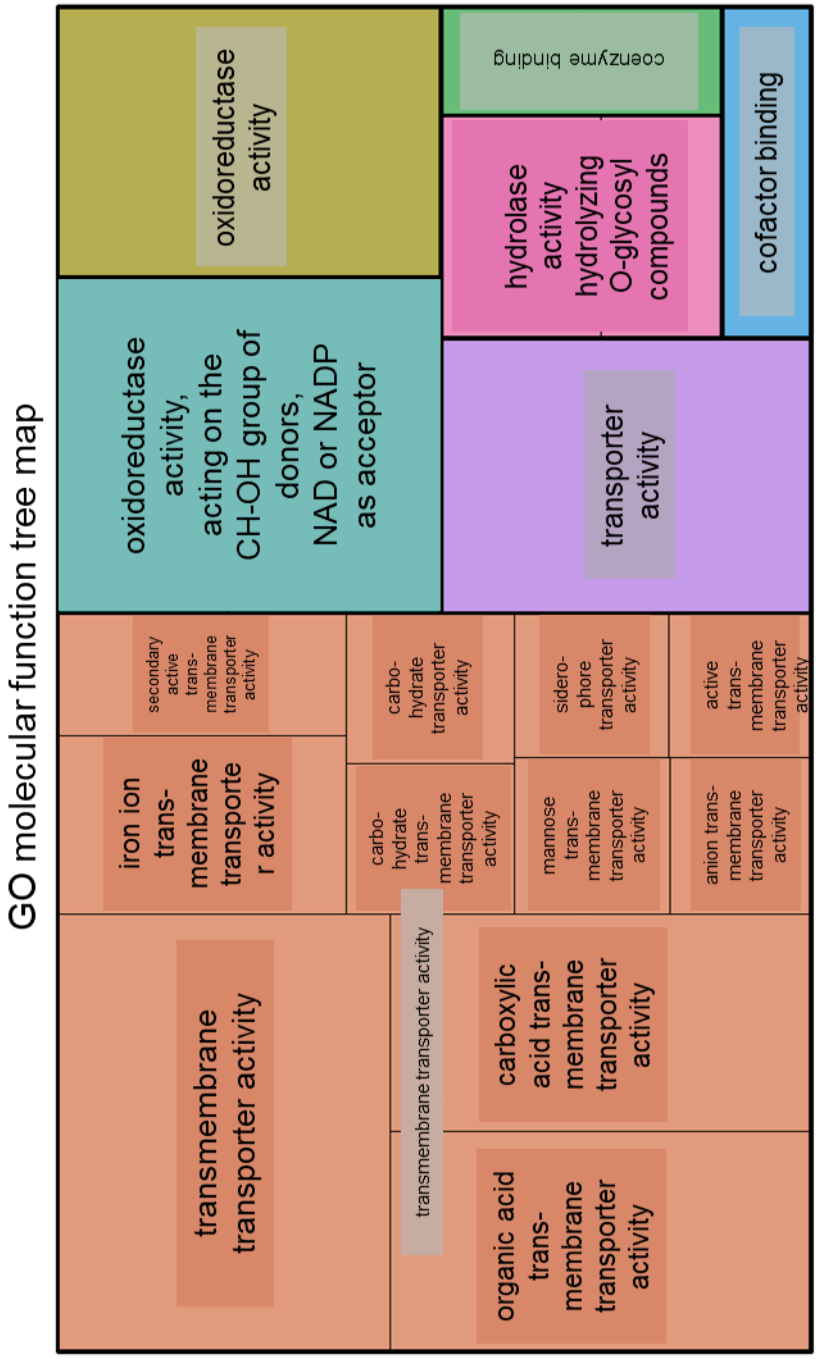
Supplementary figure S2 - Protein RNA and cell volume distributions trends along growth. Facs distribution from cells sampled at T1,T2 and T3 (see figure 1A). Up: cells on ammonium sulphate. Down: cells on glutamate.

T1	T2	T3	Name	GO Biol Proc	
			TAT2	aromatic amino acid transport	transport
			MMP1	S-methylmethionine transport	
			UGA4	gamma-aminobutyric acid transport	
			MCH2	transport	
			GIT1	glycerol-3-phosphate transport	
			PDR12	organic acid transport	
			COS12	protein targeting to vacuole	
			FET3	high-affinity iron ion transmembrane transport	
			FIT2	siderophore transport	
			ARO9	aromatic amino acid family metabolic process	Ehrlich pathway
			ARO10	aromatic amino acid family catabolic process to alcohol via Ehrlich pathway	
			INO1	inositol biosynthetic process	fatty acid catabolism
			YSR3	sphingolipid biosynthetic process	
			ICL2	propionate breakdown	
			TOS3	glucose metabolic process	gluconeogenesis
			GTO1	glutathione metabolic process	
			SPL2	protein targeting to vacuole	other processes
			HOP1	meiotic recombination checkpoint	
			YBR056W		unknown
			YKL107W		
			YDL241W		
			YGL262W		

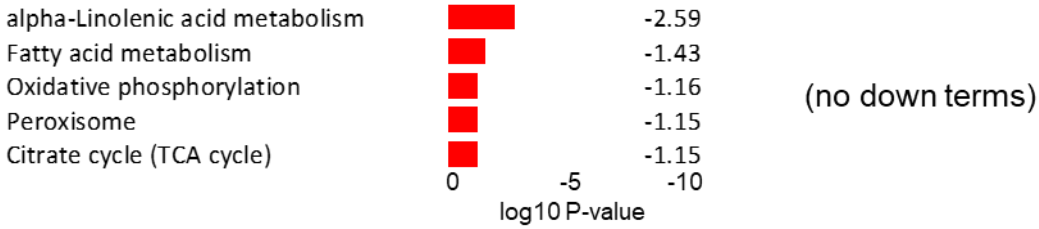
Supplementary figure S1 – GO (Biological process) classification of 22 core DEGS



Supplementary figure S3 - REVIGO ANALYSIS of Gene Ontology enrichment. Gene Ontology biological process enrichment of up- and down-regulated genes between cells growing on glutamate vs. cells growing on ammonium sulphate in the 3 growth time points.

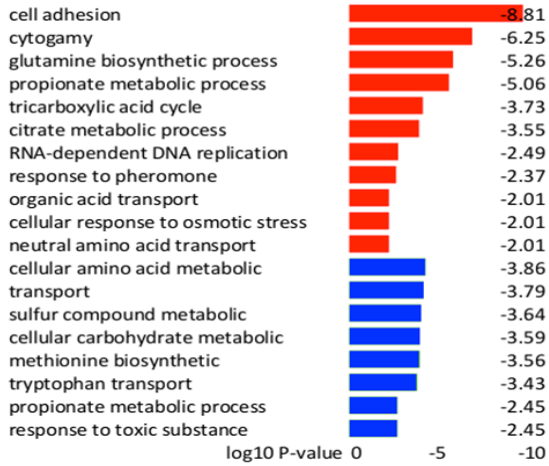


Supplementary figure S4 - REVIGO ANALYSIS of Gene Ontology enrichment. Gene Ontology molecular function enrichment of up- and down-regulated genes between cells growing on glutamate vs. cells growing on ammonium sulphate in the 3 growth time points.

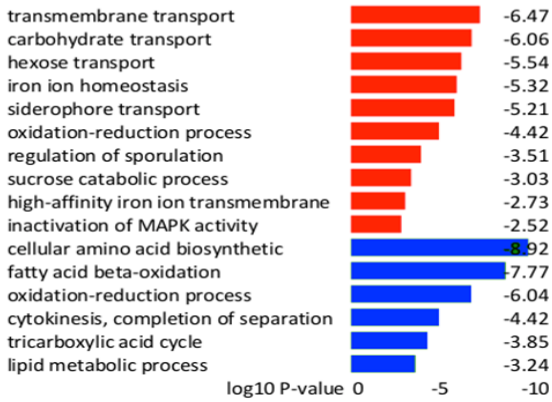


Supplementary figure S5 - Intersection T2-T3 Up and Down regulated genes – KEGG enrichment

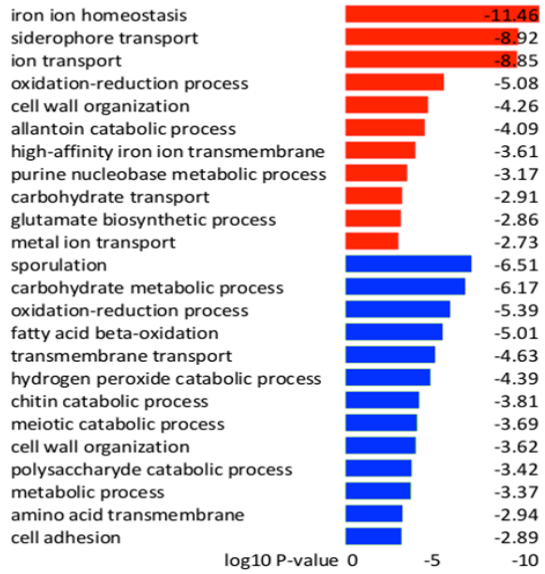
T1 Up and Down regulated genes
Biological processes



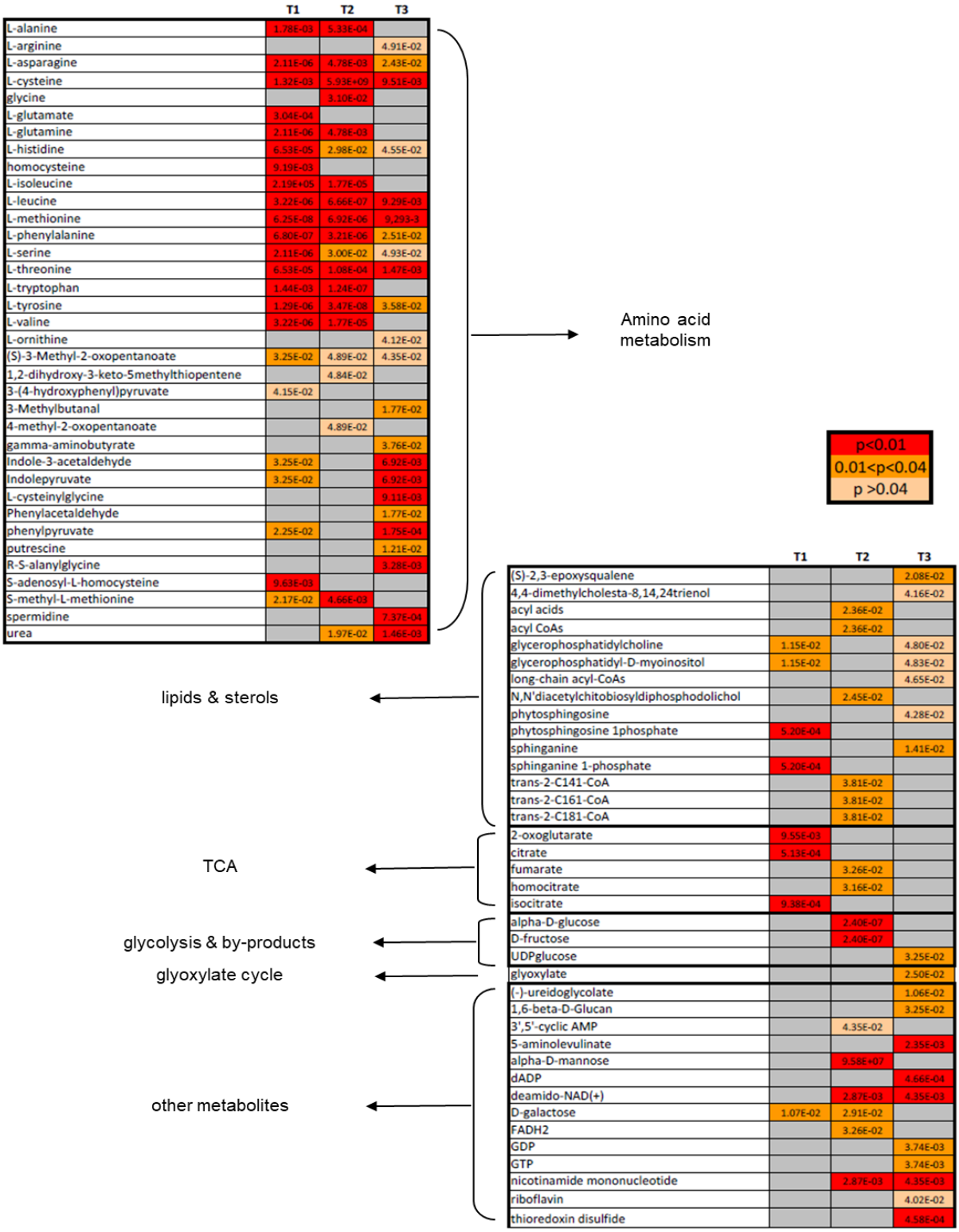
T2 Up and Down regulated genes
Biological processes



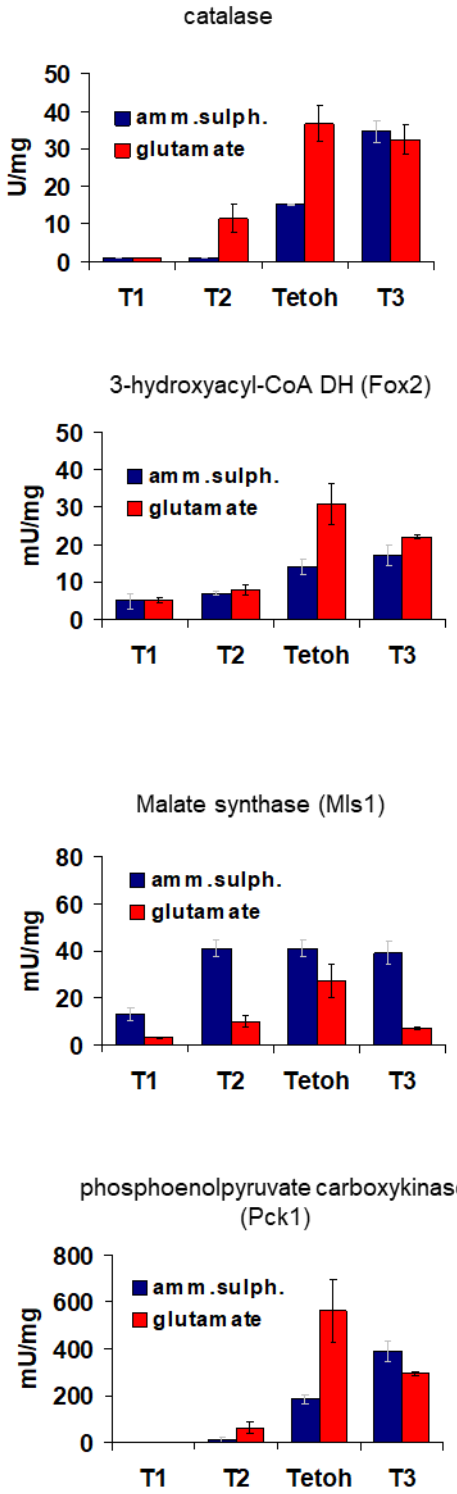
T3 Up and Down regulated genes
Biological processes



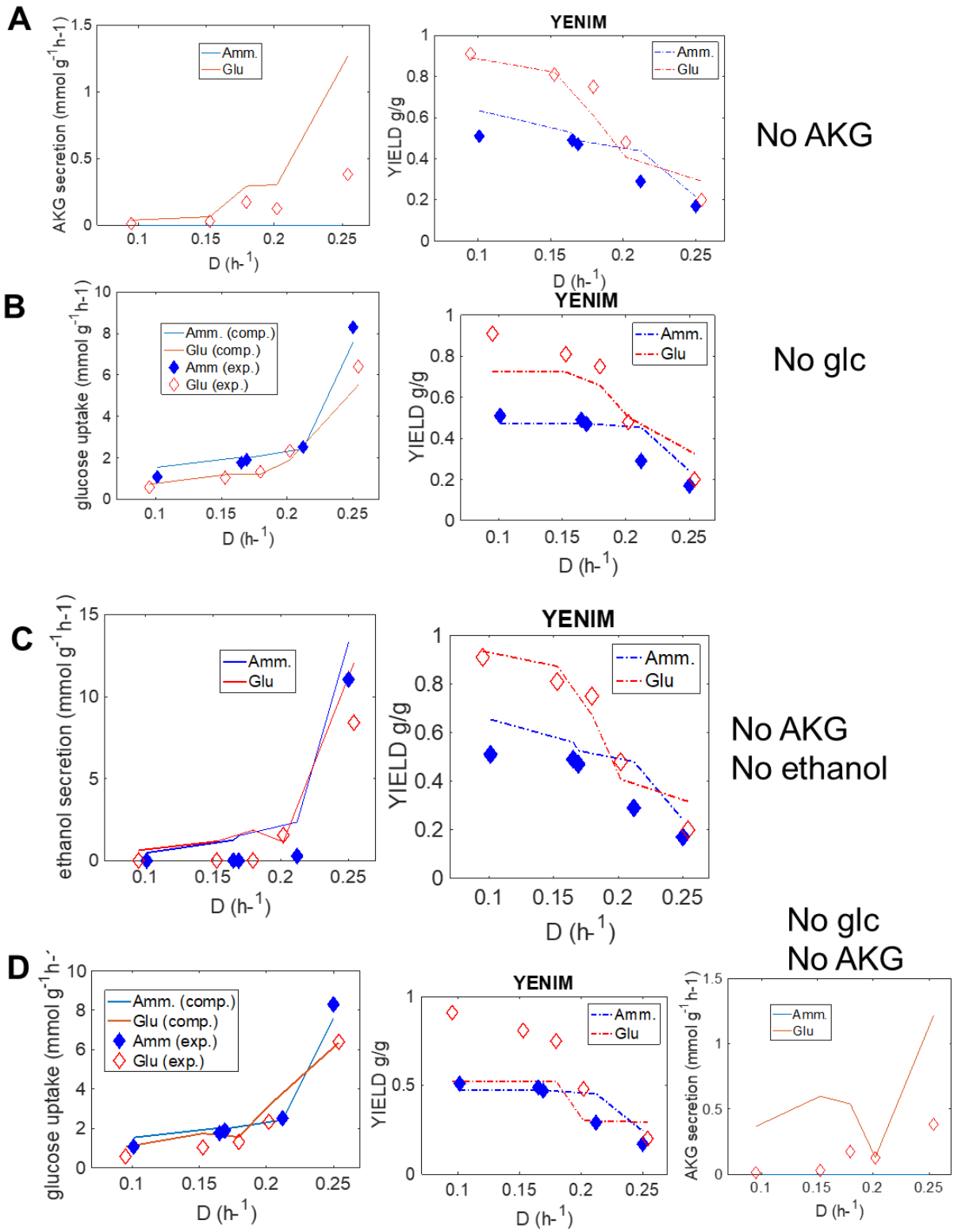
Supplementary figure S6 - Gene Ontology enrichment of up- and down-regulated genes between cells growing on glutamate vs. cells growing on ammonium sulphate in the 3 growth time points.



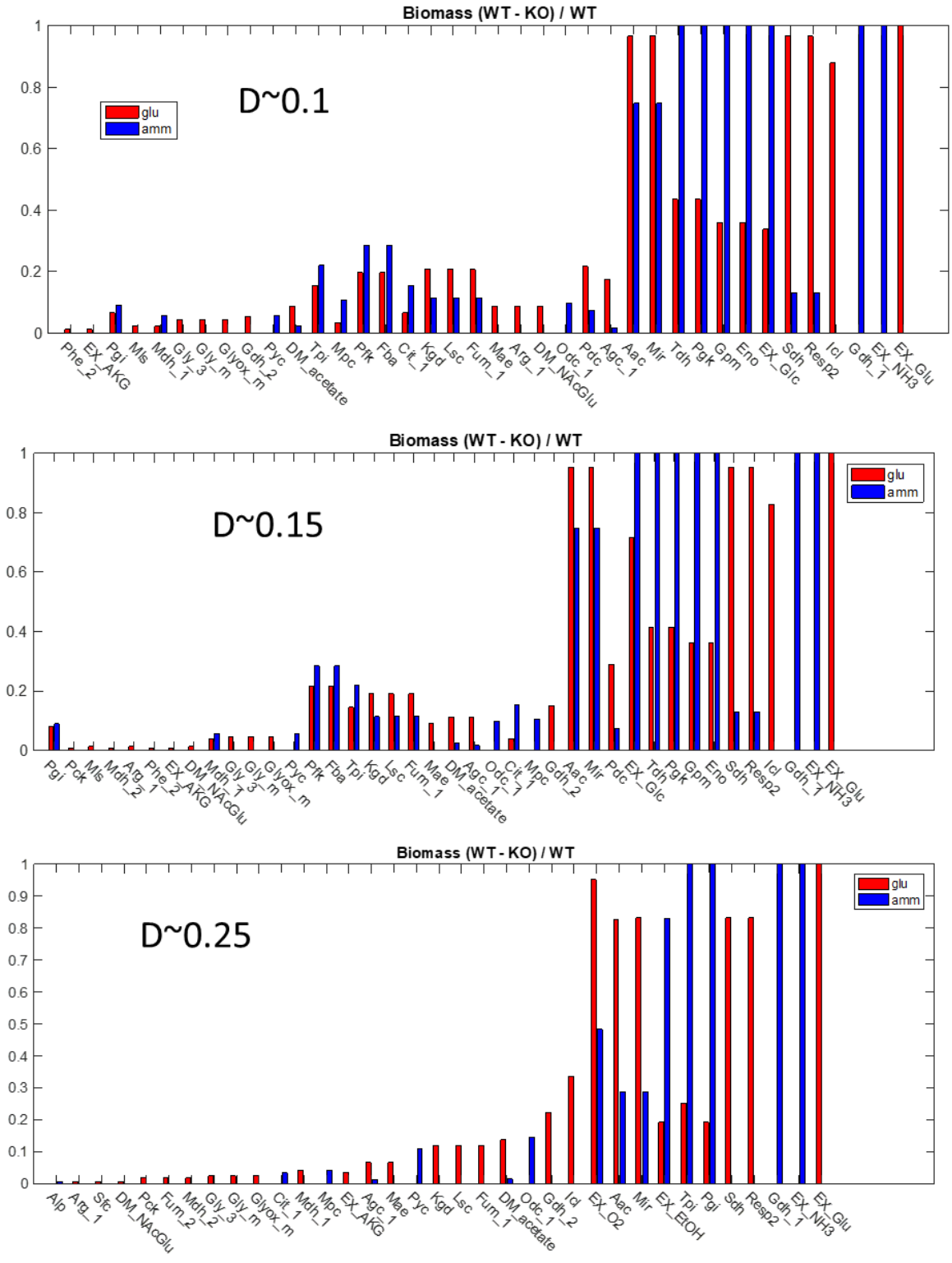
Supplementary figure S7 - List and categorization of reporter metabolites retrieved from the transcriptomics data



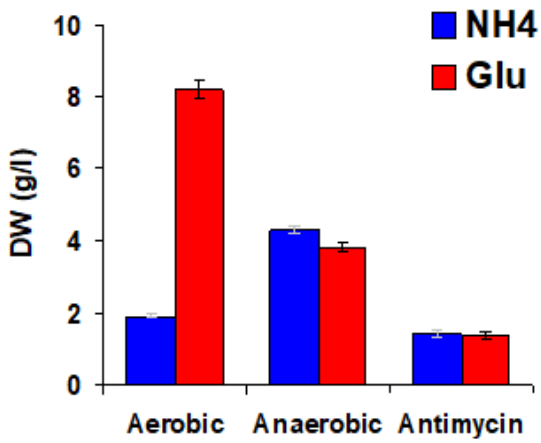
Supplementary figure S8 - Enzymatic activities of interesting enzymes (reported in Figure 4 as fold



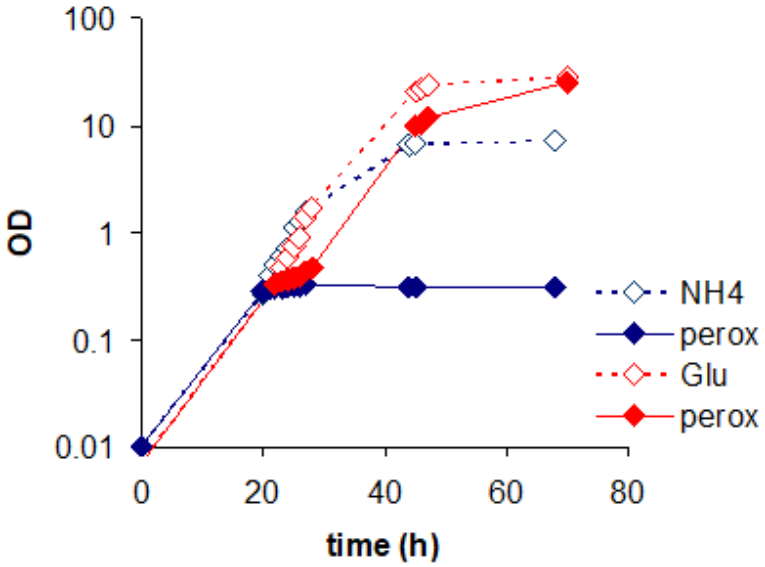
Supplementary figure S9 - Computational simulation of biomass yield of glutamate and ammonium cultures at different dilution rates.



Supplementary figure S10 – Reaction deletion analysis for different dilution rates



Supplementary figure S12 – Final biomass accumulation for ammonum and glutamate cultures in aerobic and anaerobic condition and in the presence of the



Supplementary figure S11 - Minimum amounts (0.75 mM) of H_2O_2 were added (arrow) during the exponential phase to cells growing on standard synthetic media supplemented with ammonium sulphate (blue diamonds) or glutamate (red diamonds). While in the first medium growth was completely abolished, on glutamate the treatment resulted only in a delay, and cells were able to recover and to reach the usual density (empty symbols: untreated controls).

Chapter 2: Alterations of the nutrient sensing pathways allow yeast to utilize glutamate as a combined carbon and nitrogen source

Brambilla L[1,3]*, Gnugnoli M[1,3]*, Damiani C[2,3], Porro D[1,3], Frascotti G[1,3], Alberghina L[1,3], Vanoni M[1,3]

1. Department of Biotechnology and Biosciences, University of Milano-Bicocca
2. Department of Informatics, Systems and Communication, University of Milano-Bicocca
3. SYSBIO.it - Centre of Systems Biology, Milan

*Equal contributors.

Abstract

Saccharomyces cerevisiae is able to grow on different nitrogen sources. Previous data from our laboratory showed that growth on glutamate causes an increase in biomass due to the utilization of glutamate carbon atoms for biosynthetic purposes. Moreover, simulations using a Flux Balance Analysis (FBA) model of yeast metabolism suggest that growth on media containing glutamate as the only carbon and nitrogen source is possible. Despite these considerations, a wild-type *S. cerevisiae* strain is not able to grow in these conditions. To understand the reasons underlying this growth inability, we applied a laboratory evolution approach and we isolated 4 different mutants able to grow in the presence of just glutamate, vitamins and supplements.

Sequencing of the mutant clones showed that they all carry mutations affecting the Ras/cAMP/PKA pathway. These results show that modification of the carbon sensing pathways can alter metabolism, allowing yeast to utilize glutamate in a way that the wild type strain cannot.

Transcriptomics analysis of the more interesting mutant revealed a general up-regulation of the biosynthesis of amino acids and nucleotides, as well as an enhanced expression of plasma membrane transporter genes, including glutamate permeases. Together with the enhancement of energy producing pathways, like fatty acids β-oxidation, we propose that these changes are among the driving forces in the adaptation of the evolved mutants.

Introduction

Microorganisms react to being exposed to different environments and nutrients by activating different metabolic and signalling pathways, that enable them to make the best out of the situation they find themselves in (Zaman *et al.*, 2008). Evolution and natural selection shape the cellular metabolism and its regulation. As a result, the metabolic and signalling network can sustain growth on different nutrients and can do that with different efficiency.

Glutamate is a preferred nitrogen source for *Saccharomyces cerevisiae*, i.e. it is able to sustain growth at a high rate and activate the Nitrogen Catabolite Repression (NCR) to shut down the pathways required for the utilization of alternative nitrogen sources (Godard *et al.*, 2007). The contribution of glutamate to the anabolism of nitrogen-containing compounds within the yeast cell is remarkable, as 85% of the total nitrogen groups used for this purpose originates from this amino acid.

After being internalized into the cell, glutamate is mostly degraded by the action of glutamate dehydrogenase Gdh2, which breaks it into ammonium and alpha-ketoglutarate (AKG) (Cooper, 1982). The latter can be used to fuel the TCA cycle. The presence of this carbon backbone readily available, suggest a possible C-anabolic use of the molecule.

As a matter of fact, several yeast species, like *Scheffersomyces stipites* and *Candida shehatae,* are capable of utilizing glutamate, as well as other amino acids, as a combined carbon and nitrogen source (Freese *et al.*, 2011) and the homologue of budding yeast Gdh2 is required for this metabolism.

We recently uncovered the molecular changes induced by the use of glutamate as a nitrogen source in the yeast *Saccharomyces cerevisiae*. In particular, we elucidated the transcriptional and metabolic rewiring that allows it to incorporate the carbon backbone of glutamate and exploit it for biosynthetic purposes. We found an interaction between the carbon and the nitrogen source and an accumulation of reserve compounds in cell growing in glutamate (Brambilla et al., in preparation).

Despite these considerations, wild type *S. cerevisiae* fails to grow using glutamate as a carbon source. The aim of this work is to understand the reason for this growth inability, to discriminate if it is purely a matter of metabolic network or if other cellular components are in play.

To gain further insights on the underlying molecular reason, we used both computational and wet lab analyses. In particular, we exploited laboratory evolution to isolate mutants which were able to use glutamate as carbon and nitrogen source.

Here we present the results obtained by the selection and characterization of adaptively evolved mutants, derived from a prototrophic budding yeast strain, which are indeed capable of using the backbone of glutamate as a carbon source.

The integration of physiological, computational and genomic analysis allowed us to affirm that *S. cerevisiae* metabolism, although not being the best choice for the C-catabolic utilization of glutamate, is not completely incapable of performing this task. In fact, we found that the reason for the lack of growth for wild-type strains in this condition is due to signal transduction regulation and in particular to PKA signalling. Transcriptomics analysis of one of the evolved mutant pointed hints at the regulation of biosynthesis and transporters as a possible explanation for the mutants behaviour.

The recurrent appearance of signal transduction mutants in different laboratory evolution experiments (Long *et al.*, 2015) (Li *et al.*, 2018) (Hong *et al.*, 2011) will be discussed.

Materials and Methods

Strains and media

The prototrophic *S. cerevisiae* strain GRFc (Brambilla *et al.*, 1999) was used as a reference strain and was used to perform adaptive evolution. Inoculum cultures were prepared from −80°C glycerol stocks on YPD agar plates containing 10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose and 20 g/L agar.

Media were prepared according to (Verduyn *et al.*, 1992) with 14 g/L glutamate monosodium salt monohydrate as nitrogen source and glucose or ethanol at different concentrations when an additional carbon source was needed.

Shake flask cultures

Precultured cells were inoculated in 500 mL Erlenmeyer flasks containing 100 ml of minimal medium (Verduyn *et al.*, 1992) with the same composition of cultures. The starting OD_{660}, ranging between 0.02 and 0.2, was suitable to reach the desired growth phase the day of sampling.

The cultures were grown at 30°C in a rotary shaker at 150 rpm.

Batch cultivation in fermenter

Cultivations in fermenter were run in Biostat-B fermenters (B-Braun), filled with 1.5 L of defined medium with vitamins and trace metals (Verduyn *et al.*, 1992). Cultures were grown at 30° C at pH 5.0 and 0.2 M KOH was used as a titrant.

Glucose-limited aerobic chemostat cultivations

Chemostat cultivation was performed in Biostat-B fermentors (B-Braun). A defined medium with vitamins and trace metals was used (Verduyn *et al.*, 1992). The glucose concentration in the reservoir medium was 8 g/L. A constant working volume of 1300 mL was maintained via an effluent line coupled to a peristaltic pump. A dissolved oxygen concentration above 30% of air saturation was maintained by an air flow of 1.3/min (1 v/v/m) and a stirrer speed of 1000 rpm. The temperature was maintained at 30°C and the culture pH was kept at 5.0 by automatic addition of 2 M KOH.

Cultures were assumed to be in steady state when at least six volume changes had passed since the last change in growth conditions and the culture did not exhibit metabolic oscillations.

Measurement of growth parameters

Optical density of the culture was measured at 660 nm using an Ultrospec 500 pro spectrophotometer (Amersham).

Adaptive evolution

The reference strain was grown in shaking flasks containing the described medium with 0.2 g/L of glucose as a carbon source and glutamate as a nitrogen source. Cultures were serially diluted into fresh growth medium and, after ≈ 80 generations, cells were plated onto agar plates containing only glutamate as carbon and nitrogen source. Cells were then inoculated in shaking flasks containing the same medium as the plates. The cultures were again serially diluted and after ≈ 30 generations they were plated and clones were tested again to confirm their ability to grow.

Genome sequencing

The genome of the four most interesting clones, along with the reference strain, were sequenced with the next-generation Illumina technology (IGA technology services) to the aim of identifying differences within the open reading frames, relative to the wild-type reference strain.

RNA preparation and hybridization

Cells were collected by filtration at the indicated time points and total RNA was extracted with the phenol-chloroform method (Schmitt, Brown and Trumpower, 1990) from biological triplicate samples in the and analyzed using Affymetrix Genechip Yeast Genome 2.0 Arrays, containing approximately 5,744 probe sets for 5,841 of the 5,845 genes present in *S. cerevisiae*. Messenger RNA was amplified and hybridized onto the array (Biogem S.c.ar.l) as recommended by Affymetrix.

Microarray data analysis

Raw data (CEL files) were analyzed by Biogem S.c.ar.l using the R-Bioconductor software. More than 500 genes were identified as differentially expressed (DEGs) between the different culture conditions (Fold Change >2; Corrected P value <0,01). Fold changes were converted to log2 values.

Functional analysis of DEGs

To identify Gene Ontology and KEGG terms over-represented in the differentially expressed genes identified by the transcriptomic analysis the online tool FIDEA (http://circe.med.uniroma1.it/fidea/) was used to perform an enrichment analysis. The p-value threshold was set at 0.05.

To obtain further information on the regulatory network of the DEGs, the “Rank by TF” tool of the Yeastract database (http://www.yeastract.com/) was used to predict a ranking of the main transcription factors (TFs) for responsible for the regulation of our genes of interest.

Flux Balance Analysis

To run FBA simulations a stoichiometric matrix *S* and a set of constraints (upper and lower bound of fluxes) are required. The steady state constraint is defined by the equation $dx/dt = S \cdot v = 0$, where dx/dt are time derivatives of metabolite concentrations represented by the product of the $m \times n$ matrix *S* times the vector of fluxes $v = (v_1, v_2, \ldots, v_n)$, where v_i is the flux of reaction *i*, *n* is the number of reactions, and *m* is the number of metabolites. The ensemble of functional states that the system can reach given a boundary condition *I* determines the feasible solutions space $\Phi = \Sigma \cap I$. By exploiting linear programming, FBA allows for optimization of the flux through a weighted sum of fluxes. In particular, we used the COBRA Toolbox and the GLPK solver. Because FBA only returns a single solution, we exploited Flux Variability Analysis to assess the flux variation range consistent with the maintenance of the maximal growth obtained with FBA (Damiani *et al.*, 2017).

Determination of stress resistance

Exponentially growing cells in synthetic medium were collected at an optical density between 0.3-0.5, concentrated and exposed to the following stresses: 51°C (10 min); sorbitol 4M (1 h); LiCl 2M (3h); H_2O_2 2mM (1h). Serial dilutions of the suspension were plated on YPD plates and incubated for 3 days at 30°C.

Results

Computational simulations predict growth on glutamate as carbon and nitrogen source

We previously showed (Brambilla et al, in preparation) that the carbon backbone of the glutamate molecule can be incorporated by yeast during growth the post-diauxic growth phase on ethanol and allows an enhanced growth of *S. cerevisiae*, compared with a standard nitrogen source like ammonium sulphate.

However, unlike other yeasts, S. cerevisiae is not able to grow on media containing only glutamate as carbon and nitrogen source (Freese *et al.*, 2011).

To investigate this apparent contradiction, we took advantage of a core model of yeast metabolism, which we previously developed (Frascotti et al, in preparation). We performed constraint-based simulations of yeast growth in the presence of glutamate and either glucose, ethanol or none additional carbon source. Interestingly, we could observe a theoretical growth in the presence of glutamate as the only carbon and nitrogen source (Figure 1).

This result suggests that budding yeast metabolism, and in particular the reactions included in our core model could support the use of glutamate to provide all the carbon and nitrogen necessary for biomass production.

Therefore, the lack of growth observed by the wild-type strain could be the result of the regulation of cellular growth in the presence of such a poor carbon source.

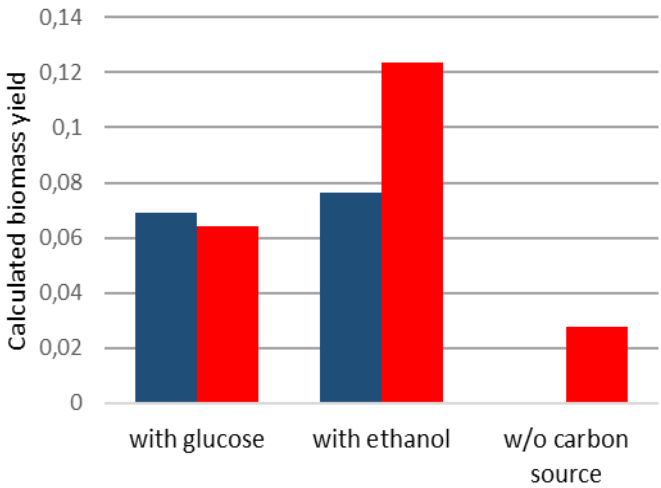


Figure 1 - Computational investigation of growth in glutamate as exclusive carbon and nitrogen source. Optimal growth yield (biomass over carbon source. Glutamate is considered a carbon source only in the absence of glucose/ethanol) as predicted by FBA - for YENIM core model

Hyperactivation of the PKA pathway allows the use of glutamate as carbon source

We decided to isolate mutants that could be able to exploit the metabolic potential highlighted by our computational simulations and, to this aim we performed an adaptive evolution starting from a prototrophic yeast strain (Figure 2A). We performed a first round of selection in a medium containing glutamate and low glucose (≈ 80 generations), followed by a second round in a minimal medium containing only glutamate as carbon and nitrogen source (≈ 30 generations).

At the end of the adaptive evolution protocol we were able to isolate 4 clones (6A, 13C, LB2, LB7) with the ability to grow using glutamate as a carbon source (Figure 2B). Analysis of the growth of the four mutants revealed differences in their growth rate and in their final biomass yield. However, all the clones grow with very high duplication times, ranging from 29 to 47 hours.

To understand the molecular reason behind the growth on glutamate, we sequenced the genome of the evolved mutants and compared it with the reference wild-type strain GRFc. As expected for growth in such a poor condition, the mutants contain many mutations throughout the genome. It was particularly useful to compare the mutations in the different strains. Strikingly, the common target of the mutations in all the four strains is the Protein kinase A (PKA) pathway. In particular, 3 out of 4 mutants carry mutations in the *GPB2* gene and the other mutant is mutated in the *IRA2* gene (Figure 2C).

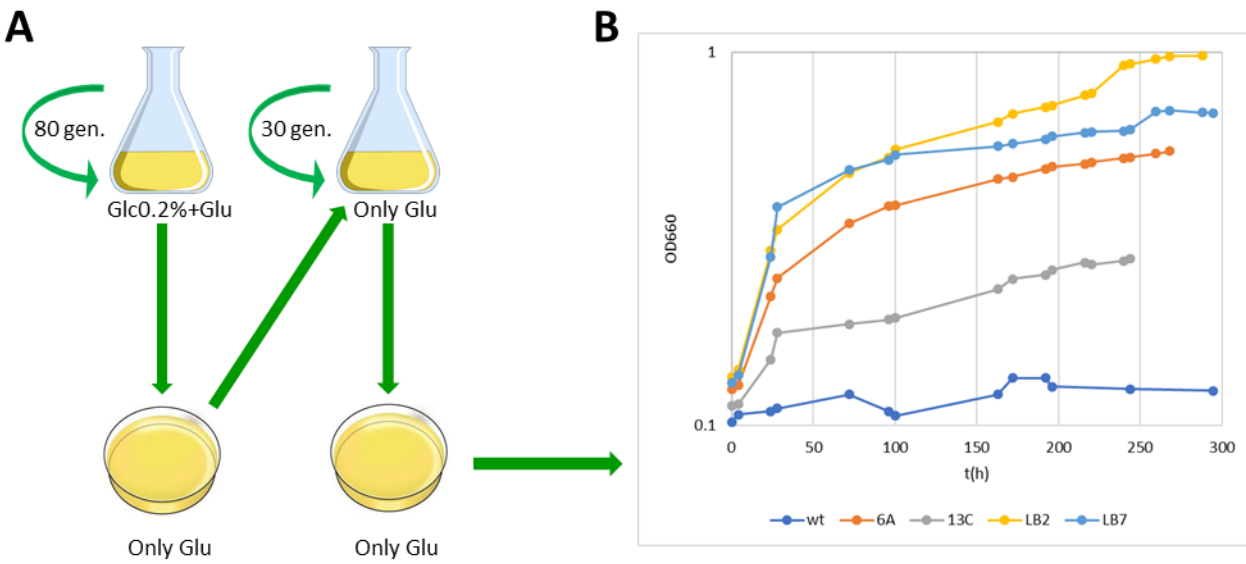
Ira2 is one of the two GTPase activating proteins (GAPs) of yeast Ras proteins (Tanaka *et al.*, 1990), which enhance the fraction of GDP-bound Ras and therefore decrease Ras/cAMP/PKA signalling (Colombo *et al.*, 2004). Gpb2, together with Gpb1 is a kelch repeat protein, which is a multistep regulator of PKA pathway. They act both upstream of Ras by binding Ira1/2 C-terminal domain and stabilizing the GAPs (Harashima *et al.*, 2006) and directly on PKA, favouring the binding of the catalytic Tpk subunits to the regulatory subunit Bcy1 (Peeters *et al.*, 2006).

LB7, LB2 and 6A mutant carry nonsense mutations in the *GPB2* gene, which generate truncated versions of the protein of 190 (LB2, LB7) and 488 (6A) amino acids. Gpb2 contains 7 kelch repeats, which fold as a β-propeller and seems to be implicated in protein-protein interaction (Harashima and Heitman, 2002). In the LB7 and LB2 mutants none of the kelch repeats is present, while in the 6A mutant only 3 out of 7 repeats are present. In both cases, there is a clear alteration of its protein-protein interaction capability. On the other hand, the 13C clone carries a mutation that introduces a stop codon in the position 1952 of *IRA2*. The truncated protein generated

this way still possesses the Ras-interacting domain (or most of it) but lacks the C-terminal Gpb-interacting domain.

As a result, we predict that the mutants should have an hyperactivation of the PKA pathway. An enhanced PKA activity is known to increase sensitivity to different stresses. Consistently, both LB7 (mutated in *GPB2*) and 13C (mutated in *IRA2*) clones display a lower resistance to temperature, oxidative, osmotic and saline stresses, compared to the wild-type strain (Figure 2D).

To further strengthen the causal link between the mutations and the growth phenotype, we deleted the entire *GPB2* gene and observed that this deletion allows yeast to grow on glutamate medium, even if the phenotype is worse than that one of the GPB2-mutants, LB7. (Supplementary figure S1).



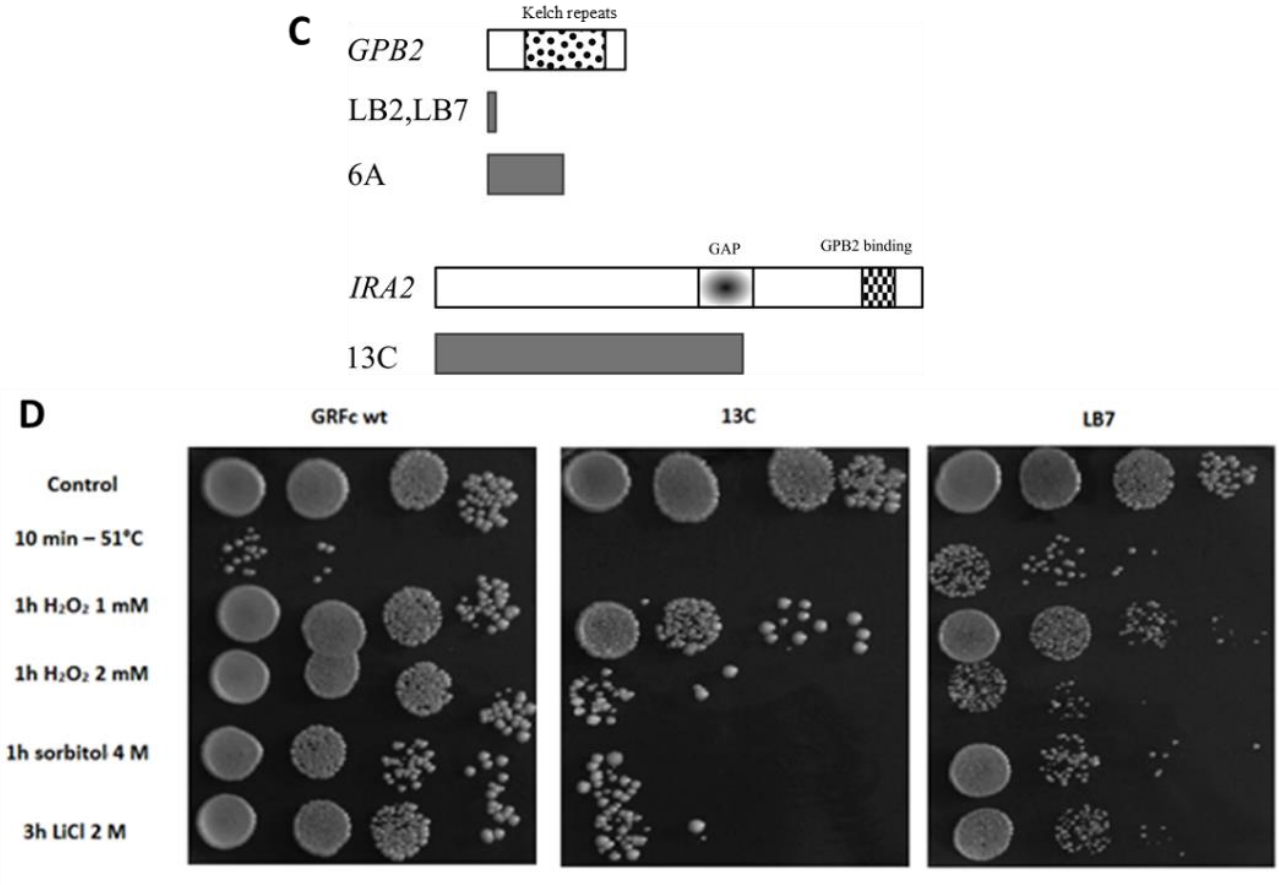


Figure 2 – Laboratory evolution and chharacterization of the evolved mutant capable of growing on glutamate as a carbon source. (A) Scheme of the performed laboratory evolution. (B) Growth of the four isolated mutants in a mineral medium containing only glutamate, vitamins and essential supplements. (C) Visualization of the most interesting mutations present in the four mutants and their relationship with the protein domains. (D) Drop-test for stress resistance. Aliquots (0.5 OD) of exponentially growing cells (in minimal medium with glucose) have been collected and exposed to the following stresses: hydrogen peroxide 1 or 2 mM (1 hr), 50°C (10 minutes), LiCl 2M (3 hr), Sorbitol 4 M (1 hr). Sequential dilution of treated and untreated cells were spotted on YPD and incubated for 3 days.

C-assimilation of glutamate is enhanced by pH and possibly oxygenation control

For further analyses, we chose to use the mutant which showed the best growth phenotype, which was LB7. To assess the importance of different physiological parameters for the growth on glutamate only medium, we cultivated the strain in batch fermentor, keeping the pH at 5 and the dO_2 above 50%. In this condition, LB7 strain still grows with slow growth rate, but, unlike the shaking flask experiments, it is able to reach a much higher biomass, up to 10 OD_{660}. Moreover, it is able to consume all the glutamate that is contained in the medium (Figure 3A).

To evaluate the impact of the different parameters for LB7 growth, we performed the same growths but without control of oxygenation and pH. LB7 grown in this condition reaches a final biomass comparable to that of the growth in the shaking flasks. We also monitored the parameters during this growth and found out that, while the pO_2 stays constant for all the growth, the pH changes from an initial value below 6 to almost 7. These results suggest a greater importance of pH control rather than oxygenation control for LB7 growth. Consistently, mutant cells (LB7 and *gpb2Δ*), when grown in flasks with 100 mM MES buffer at pH 5.5 the final biomass accumulation showed an increase of more than 3 folds, even if not comparable to that of the fermentor growths.

For this last reason, we think that oxygen limitation could also be a factor in the difficult growth of the mutant cells using glutamate as carbon and nitrogen source.

Indeed, growth in glucose-limited chemostat allowed us to evaluate the respiratory parameters (O_2 consumption and CO_2 production) of LB7, which are enhanced in this condition. Moreover, it allowed us to measure the biomass production, which is reduced compared to the wild-type strain (Figure 3B). This is consistent with literature data that indicate a trade-off in mutant isolated from laboratory evolution, which result less efficient than the wild-type in rich media (Wenger *et al.*, 2011) (Hong and Nielsen, 2013).

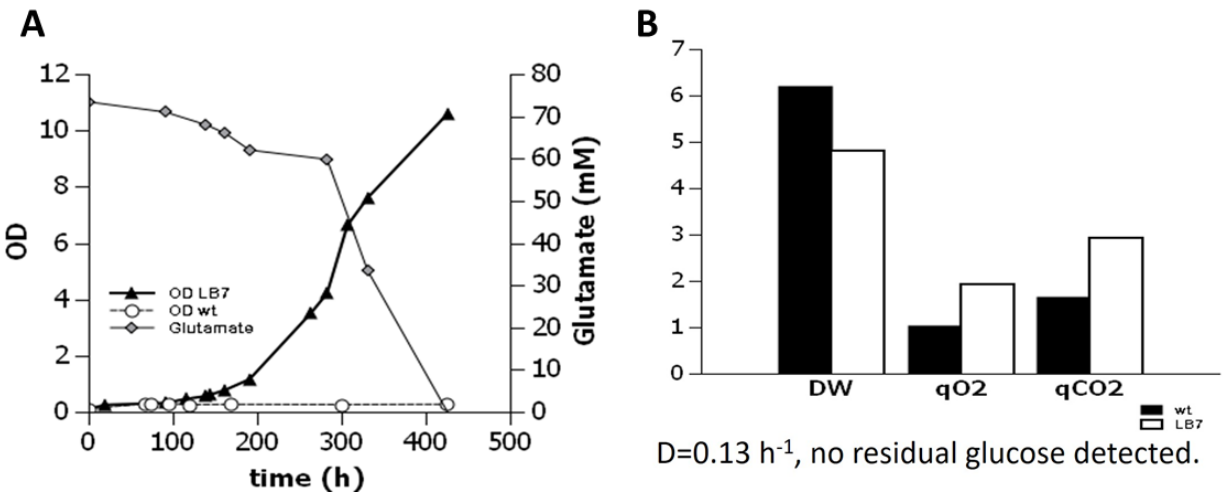


Figure 3 - Evaluation of LB7 properties in fermentor and chemostat. (A) LB7 cells were grown in fermentor at controlled pH and oxygenation. Growth was evaluated by spectrophotometry and external glutamate concentration was measure by NMR. (B) Physiological parameters of LB7 (white) and wild-type (black) growing in a glucose-limited chemostat. DW=dry weight; qO2=oxygen consumption; qCO2=carbon dioxide production.

Transcriptomic analysis shows a modulation of metabolism and transport in LB7 cells

To have a more comprehensive view of the changes occurring into the mutant cells that could allow growth on glutamate as a carbon source, we investigated the transcriptional changes that occurred in such condition. Due to the inability of the wild-type cells to use glutamate as a carbon source, we lacked a proper control for our analysis. Therefore, we decided to investigate the changes in transcription caused by the mutation and those caused by the growth in glutamate separately (Figure 4A).

To this aim, we cultivated wild-type GRFc cells in fermentor, in a medium containing glutamate as a nitrogen source and ethanol as a carbon source, to allow respiratory growth and to avoid glucose repression (condition A). Our condition B was the same growth performed using the LB7 strain. Finally, we cultivated LB7 cells in a medium containing only glutamate as carbon and nitrogen source (condition C).

We extracted the RNA and analysed the transcriptome using Affymetrix Yeast 2.0 microarray. Strikingly, as shown in the dendrogram of our samples, the C condition is transcriptionally more similar to the A condition than to the B (Figure 4).

With the data obtained we performed two comparisons: B vs A, which told us which genes were differentially expressed between wild-type and mutant cells in the presence of glutamate and ethanol, and C vs B, which showed us the changes in expression due to the growth in our unusual growth medium.

As shown in Figure 4C, the widest transcriptional rearrangement occurs in the B vs A comparison, where we found more than 600 differentially expressed genes (DEGs). In this comparison, the majority of the genes are up-regulated (fig). The C vs B comparison, on the other hand, shows a more modest change in transcription (≈ 200 DEGs), with a predominance of down-regulated genes. Finally, 118 genes were common between the 2 lists of DEGs, which were all regulated in opposite ways in the two comparisons (Figure 4D).

To gain further insights into the biological pathways and functions affected in the two comparisons, we used the web tool FIDEA, which is able to retrieve functional terms from Gene Ontology and KEGG databases significantly associated with a list of genes.

Figure 4E shows the terms retrieved from the Gene Ontology slim (a subset of the GO terms) significantly associated with the DEGs derived from the two comparisons under exam. Many of the terms associated with the B vs A gene list are shared with those related to the C vs B comparison. In particular, transcription of genes involved in metabolic processes such as small molecule, amino acid and sulphur compounds metabolism and transmembrane transport are affected both by LB7 genotypic changes

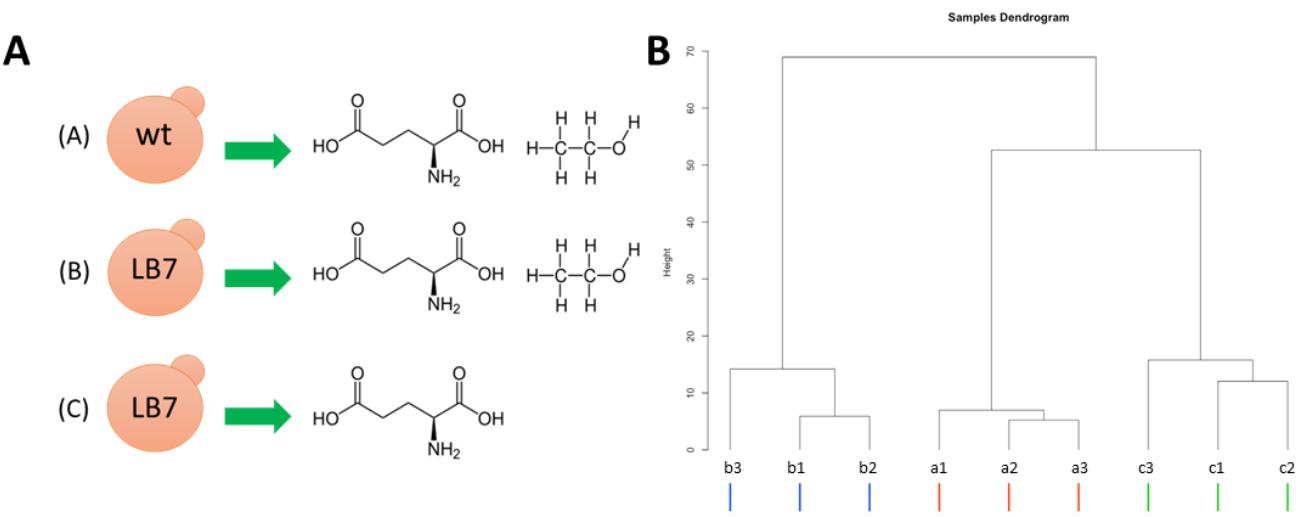
and by the growth on glutamate as a carbon and nitrogen source. Resembling what has been observed for the general analysis, all these terms are mostly up-regulated in the LB7 mutant compared to the wild-type, and down regulated when the mutant is grown in glutamate. Strikingly, the only term significantly associated with the C vs B comparison is the one related to the peroxisomal compartment.

KEGG terms enrichments (Supplementary figure S2A) confirm the results, elucidating regulation of the different amino acid biosynthetic pathways, which will be discussed in the next paragraph.

The analysis of the GO cellular compartments terms (Supplementary figure S2B) reveals, for the B vs A comparison, changes in the external compartment of the cell, with a ≈75% of up-regulated genes whose product is localized in either the cell wall or the plasma membrane. The latter term is also associated with the list of DEGs in the C vs B comparison, with a slight predominance in up-regulation. Consistent with the GO slim analysis, a wide up-regulation of genes associated to the peroxisomal matrix is observed.

Biological processes related to our DEGs are mostly metabolic, reflecting what we observe with the KEGG terms. Additionally, mutated cells compared to wild-type (B vs A, Supplementary figure S3) show a downregulation in the assembly of the respiratory chain and growth in glutamate only (C vs B, Supplementary figure S4) also down-regulates purine biosynthesis.

From this preliminary analysis we can conclude that hyperactivation of the PKA pathway induces an up-regulation of different metabolic processes (mostly biosynthetic) and of the transport of molecules in the cell, while the growth in



glutamate medium downregulates genes in the same pathways, but with a lower extent.

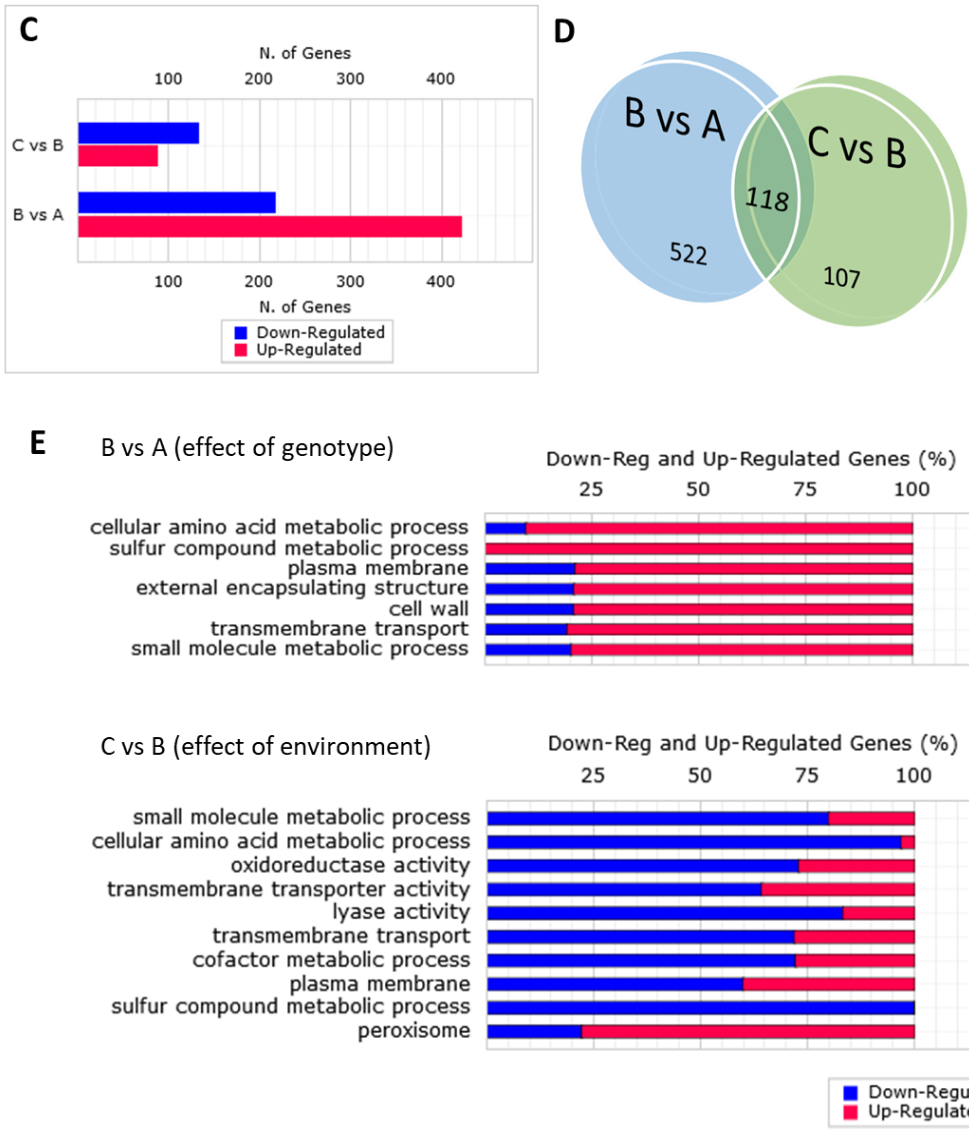


Figure 4 – (A) Scheme of the transcriptomic analysis of LB7 mutant. Condition A: wild type GRFc strain growing in ethanol+glutamate; Condition B: LB7 mutant growing on ethanol+glutamate; Condition C: LB7 mutant growing on glutamate only. (B) Correlation treemap of samples similarities. (C) Significant differentially expressed genes in the B vs A and C vs B comparisons (fold change>2, p-value<0.01) (D) Number of DEGs in the two comparisons ain intersection between the two gene lists. (E) GO slim terms significantly associated with the B vs A (top) and C vs B (bottom) DEGs lists.

LB7 mutation modulates amino acid and nucleotides biosynthesis, while growth in glutamate enhances energy producing pathways

One of the most interesting terms derived from the FIDEA analysis are those related to the metabolism of amino acid. We therefore investigated in more detail the regulation of the amino acid biosynthetic pathways. As visualized in Figure 5A, in the B vs A comparison, we can observe an up-regulation of several biosynthetic pathways. In particular, we identified genes positively regulated in the pathways that synthesize the amino acids directly from glutamate (Lys, Arg, Pro), those belonging to the pyruvate (Leu, Val, Ile) and oxaloacetate (Asn, Cys) families. Consistently, analysis of the transcription factors involved in the regulation of the B vs A comparison indicates a strong role of Gcn4 in the transcriptional rearrangement of the LB7 mutant (Supplementary table S1).

Interestingly, only a small fraction of the pathways just mentioned are down-regulated in the C vs B comparison. In particular, the Lys and Arg biosynthesis. In addition, also the Tyr biosynthesis, which were not regulated by the LB7 mutation, is down-regulated by growth using glutamate as a carbon source. Additionally, the pentose phosphate pathway is up-regulated in this condition.

Another transcription factor significantly associated with both the B vs A (Supplementary table S) and C vs B (Supplementary table S2) DEG lists is Bas1, which is involved in the purine and histidine biosynthetic pathways. In fact, the initial reactions of this pathway (those leading to IMP) are up-regulated by the LB7 mutation and down-regulated by growth in glutamate as carbon and nitrogen source, at a very similar extent (Figure 5C, right). In the downstream part of the pathway, the C vs B condition seems to be correlated with a down-regulation of AMP production, while the LB7 mutation up-regulated the reaction leading to GMP production. Additionally, in the B vs A comparison, biosynthesis of the pyrimidines is up-regulated, while one of the same components of the pathway is down-regulated by the switch from the B to the C condition (Figure 5C, left).

An interesting term that emerged from the GO enrichment analysis is the one related to the peroxisome in the C vs B analysis. We reasoned that this could mean a regulation of the β-oxidation of fatty acids. As shown in Figure 5B, growth in glutamate only, compared to growth in glutamate and ethanol, causes a transcriptional up-regulation of the *POX1, FOX2* and *POT1* genes, which allow cells to convert fatty acids to acetyl-CoA, as well as the *CTA1* catalase gene, involved in the detoxification of the hydrogen peroxide produced in the process. No regulation of the β-oxidation pathway in the LB7 mutant growing on glutamate and ethanol, compared to the wild-type strain grown in the same condition (B vs A), is observed.

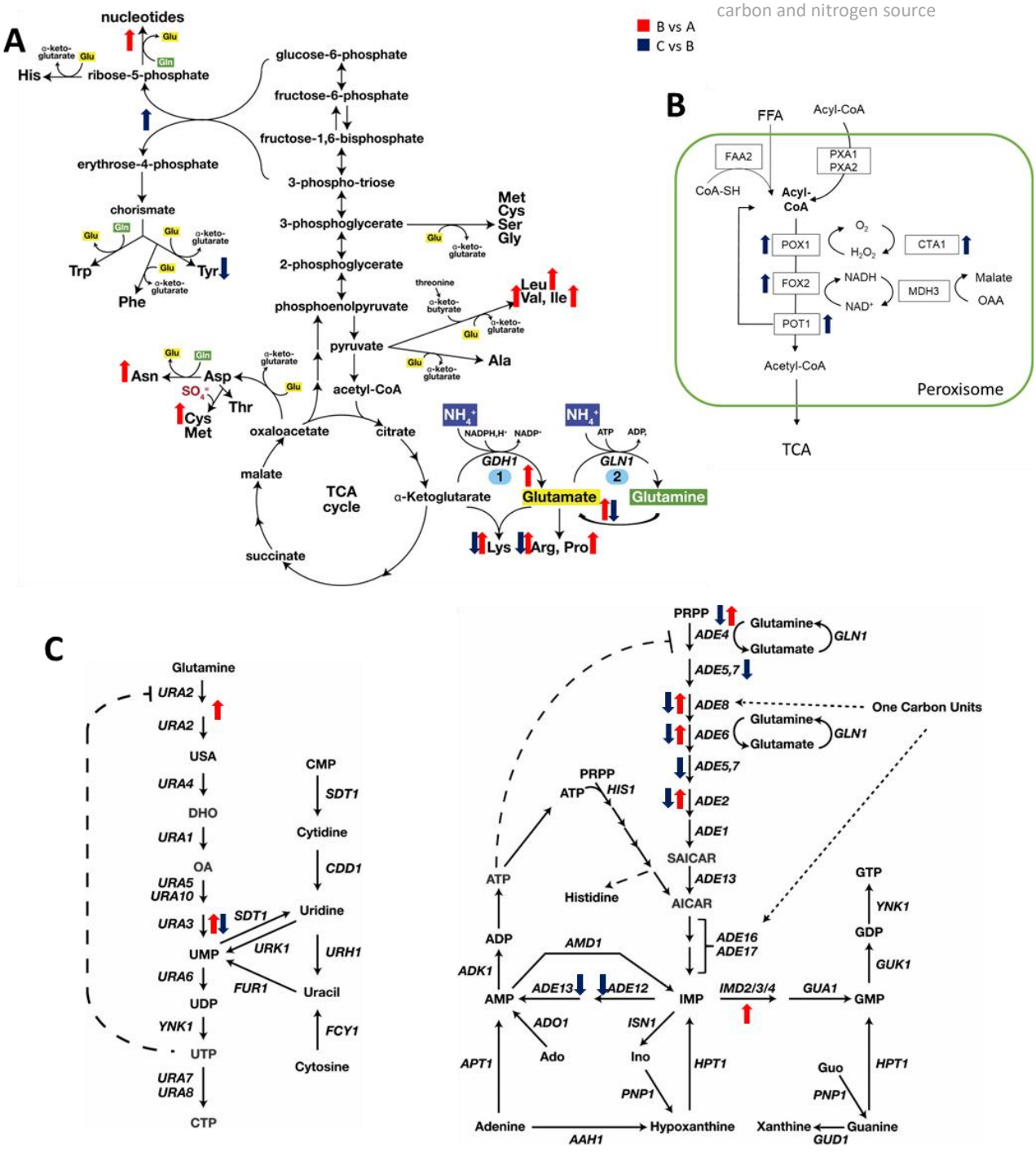


Figure 5 – Transcriptional regulation of metabolic genes involved in amino acid biosynthesis (A), fatty acid β-oxidation (B) and nucleotide biosynthesis (C, left: pyrimidine biosynthesis, right: purine and histidine biosynthesis). Regulation of transcription is represented as arrows. Upward arrows indicate up-regulation, downward arrows represent down-regulation. Red arrows are relative to the B vs A comparison, while blue arrow indicate a regulation in the C vs B comparison. Absence of arrows indicate a lack of significant alteration of the transcript levels (fold change too low or p-value too high). Figures A and B are modified from Ljungdahl and Daignan-Fornier, 2012.

LB7 shows an increase in the expression of several amino acid permeases

Finally, because of the significant association of the term related to transmembrane transport to both the comparisons, we investigated the transcriptional regulation of the genes coding for yeast transporters.

First, we obtained the list of the list of yeast transporters from the Yeast Transport Proteins Database (http://ytpdb.biopark-it.be) and we observed how many of that proteins were regulated in the two comparison we performed for the three transporters categories: plasma membrane transporters, internal membrane transporters and not yet characterized transporters (Supplementary figure S5).

About 10% of internal membrane transporters genes display a regulation both in both the comparisons performed. The same or a minor fraction of the plasma membrane transporters and uncharacterized transporters is regulated in the C vs B comparison. On the other hand, 27% of plasma membrane transporters genes is regulated by the LB7 mutation (B vs A) and, in particular, 24% of them is up-regulated. This is also true for the 16% of the still unknown transporters genes.

Because of the data just discussed, we assessed the regulation of plasma membrane transporters specifically involved in nitrogen-containing compounds transport (Figure 6). As already observed for components of their biosynthetic pathways, we found that many genes coding for amino acid permeases are up-regulated in the presence of the mutations of LB7 and the vast majority of those is not down-regulated in the C vs B comparison, which means that they still are more expressed in the LB7 strain growing on glutamate. Interestingly, among those DEGs are both the broad-range amino acid permease *GAP1* and the dicarboxylic amino acid permease *DIP5*, which are the two transporters known to be involved in glutamate import into the yeast cell. In addition to this, 5 out of 9 of the unknown transporters show similarities with amino acids and carboxylic acids permeases.

The C vs B comparison doesn't show a regulation by the culture medium of the expression of plasma membrane permeases. The only interesting data is the up-regulation of the ammonium exporter gene *ATO3*, coding for an ammonium exporter, which could be involved in a detoxification mechanism.

We can conclude that both known transporters for glutamate and putative amino acid permeases are overexpressed in the LB7 mutant, compared to the wild-type. Although growth with glutamate as the only carbon source does cause a general down-regulation in the transcription of transmembrane transporters, it does not affect the regulation of

these permeases. For this reason, it is possible to hypothesize that they their level could be high enough to play a role in the growth of LB7 in glutamate only.

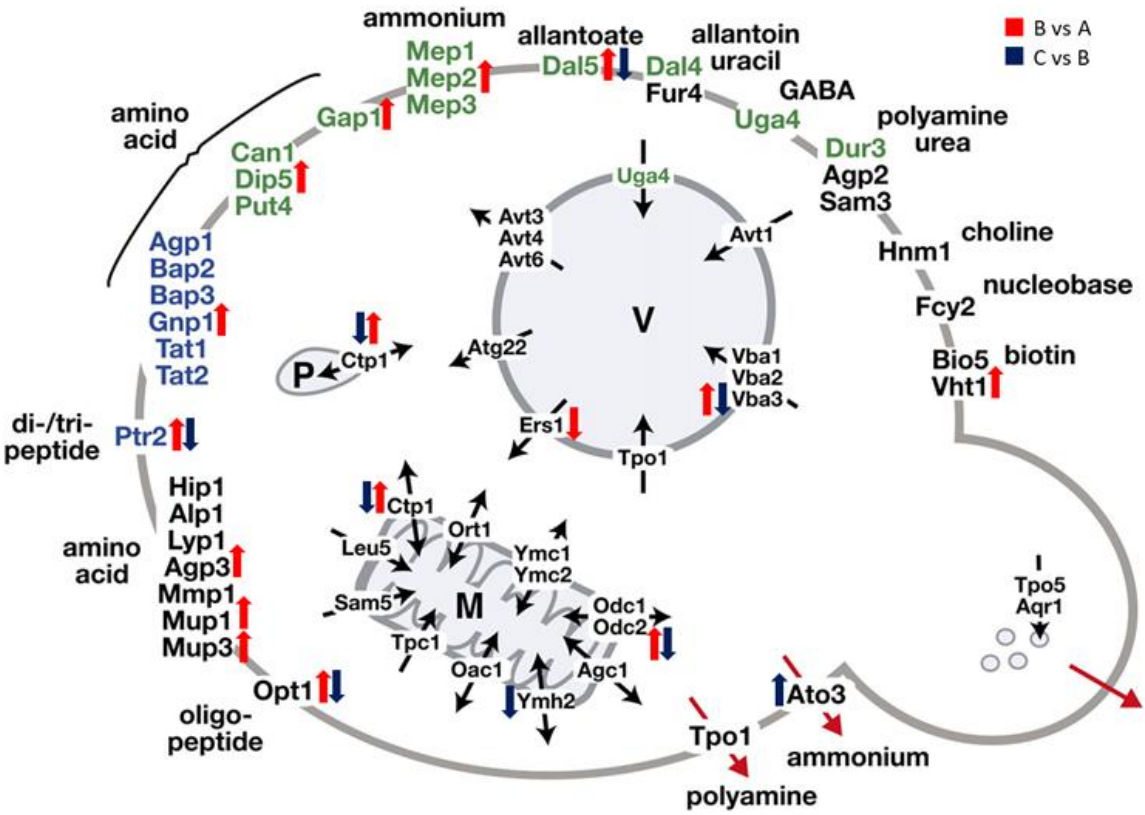


Figure 6 – Transcriptional regulation of nitrogen-related transporters (modified from Ljungdahl and Daignan-Fornier, 2012). The representation of the changes in transcription is the same as Figure 5.

Discussion

In this chapter we have studied how a prototrophic yeast strain subjected to laboratory evolution was able to acquire a completely new metabolic capability, absent in the wild-type strain: incorporate the carbon atoms of glutamate, not only as an additional source of carbon during co-consumption with ethanol, but without any other carbon compound.

The result of our computational analysis showing that yeast central metabolism can sustain growth on glutamate as carbon and nitrogen source, together with the fact that yeast is able to exploit glutamate carbon atoms during growth on ethanol (Chapter 1) indicates a regulatory mechanism behind the lack of growth of wild-type *S. cerevisiae* using glutamate as a carbon source. Considering this, it is not entirely surprising that all the four mutants we retrieved carried mutations components of the signal transduction machineries.

Nevertheless, the bioinformatic analysis of the transcriptomics results retrieved almost exclusively terms related to metabolic pathways or transport of nutrients. In particular we observe, for a fraction of terms, an up-regulation caused by the LB7 mutation (and hence, by PKA activation) and a down-regulation induced, in the mutant, by the growth in glutamate without another carbon source. Interestingly, the magnitude of the regulation in the two directions, is rarely very different in the two conditions. We have no way to evaluate the level of these transcripts in the wild-type with glutamate as the only carbon source, because it cannot sustain growth in such condition. Anyway, it is interesting to hypothesize that the down-regulation observed in the mutant reflects what happens in the wild-type. This suggest the fascinating possibility that the hyperactivation of the PKA pathway is able to trick the cell, mimicking the presence of a richer nutrients, which does not shut down all the needed pathways and can, at its slow pace, grow. This could explain the curious data that show that LB7 grown in glutamate only has a transcriptional profile more similar to that of the wild-type strain grown in glutamate and ethanol than to a LB7 grown in that condition.

Biosynthesis of amino acid and purine nucleotides seems to be central in the metabolic rewiring that is generated by the deletion of *GPB2* found in the mutant. Two of the main players involved in the transcription of genes in these pathways are Gcn4 and Bas1, whose targets are significantly enriched in the B vs A gene list. It is not the fist time that these genes have been associated with PKA. In fact, Gcn4 targets have been found up-regulated following Ras-hyperactivation (Zaman *et al.*, 2009) and the Snf1 kinase, negatively regulated by PKA, is involved in the repression of *GCN4* translation (Shirra *et al.*, 2008). Moreover, interaction between Gcn4 and Bas1 were already identified as crucial in the different Crabtree effect observed in a laboratory evolved

strain characterized by a loss of function mutation in *RAS2* (Martínez *et al.*, 2014). It is possible that their combined effect is also important for LB7 phenotype.

The effect of PKA on yeast transporters is also interesting. The LB7 strain presents a wide up-regulation of the plasma membrane transporters. The two amino acid permeases that caught our eye (*DIP5* and *GAP1*) are not usually regulated in the same way. In fact, Gap1 is under the control of the nitrogen catabolite repression, which is active when an optimal nitrogen source, like glutamate, is present (Godard *et al.*, 2007). On the other hand, Dip5 is specifically required for the import of glutamate and aspartate into the cell. It seems that the mutation is able to change the classical pattern of regulation of the two permeases. How that is done remains to be elucidated, because the effect of PKA on Gap1 so far is not absolutely clear and seems to be more inhibitory than promoting (Garrett, 2008).

In the results section we have focused mainly on the amino acid permeases, but the transcription of various sugar (mannose, fructose, glucose) transporters, sulphate and phosphate permeases and proteins devoted to the import of vitamins and other essential ions is also up-regulated in the LB7 mutant. This is consistent with the fact that mutations in components of the PKA pathway are frequent in different kinds of evolution experiments: from a generic increase of fitness compared to other competitors (Venkataram *et al.*, 2016) (Kvitek and Sherlock, 2013) to a better utilization of a non-preferred carbon source (Hong *et al.*, 2011).

In the last cited article, Nielsen and colleagues identify several strains harbouring mutations in the *RAS2* protein or in *CYR1* gene, coding for the yeast adenylate cyclase, which are able to use galactose more efficiently without having direct mutations on the pathways involved in galactose utilization. Our results strengthen the relationship between adaptation to new growth condition and the signal transduction, derived from their intrinsic pleiotropic nature.

From the point of view of applications and future perspectives, the consequences of this reasoning are different. First, laboratory evolution is confirmed to be a valuable tool for developing strains of interest carrying non-trivial mutations, which could both be directly useful for production or for better understanding growth regulation. Moreover, the great power of the modulation of the signalling pathways could be directly exploited by more focused mutagenesis, in order to obtain various mutants. Finally, it is fascinating to observe how the obtained strains perform in the different conditions, from a nutrients-rich environment (in which they usually perform worse than their reference wild-type strains) (Wenger *et al.*, 2011) to a different poor environment. Integrated analysis of the behaviour of the different strains in the presence of different nutritional sources could really boost the understanding of cellular metabolism and its regulation.

References

Brambilla, L., Bolzani, D., Compagno, C., Carrera, V., Van Dijken, J. P., Pronk, J. T., Ranzi, B. M., Alberghina, L. and Porro, D. (1999) 'NADH reoxidation does not control glycolytic flux during exposure of respiring Saccharomyces cerevisiae cultures to glucose excess', FEMS Microbiology Letters, 171(2), pp. 133–140. doi: 10.1016/S0378-1097(98)00589-8.

Colombo, S., Ronchetti, D., Thevelein, J. M., Winderickx, J. and Martegani, E. (2004) 'Activation state of the Ras2 protein and glucose-induced signaling in Saccharomyces cerevisiae', Journal of Biological Chemistry, 279(45), pp. 46715–46722. doi: 10.1074/jbc.M405136200.

Cooper, T. G. (1982) Nitrogen Metabolism in Saccharomyces cerevisiae., The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. doi: 10.1101/087969180.11B.39.

Freese, S., Vogts, T., Speer, F., Schäfer, B., Passoth, V. and Klinner, U. (2011) 'C- and N-catabolic utilization of tricarboxylic acid cycle-related amino acids by Scheffersomyces stipitis and other yeasts', Yeast. Wiley-Blackwell, 28(5), pp. 375–390. doi: 10.1002/yea.1845.

Garrett, J. M. (2008) 'Amino acid transport through the Saccharomyces cerevisiae Gap1 permease is controlled by the Ras/cAMP pathway.', The international journal of biochemistry & cell biology, 40(3), pp. 496–502. doi: 10.1016/j.biocel.2007.08.012.

Godard, P., Urrestarazu, A., Vissers, S., Kontos, K., Bontempi, G., van Helden, J. and André, B. (2007) 'Effect of 21 different nitrogen sources on global gene expression in the yeast Saccharomyces cerevisiae.', Molecular and cellular biology, 27(8), pp. 3065–3086. doi: 10.1128/MCB.01084-06.

Harashima, T., Anderson, S., Yates, J. R. and Heitman, J. (2006) 'The Kelch Proteins Gpb1 and Gpb2 Inhibit Ras Activity via Association with the Yeast RasGAP Neurofibromin Homologs Ira1 and Ira2', Molecular Cell, 22(6), pp. 819–830. doi: 10.1016/j.molcel.2006.05.011.

Harashima, T. and Heitman, J. (2002) 'The Gα protein Gpa2 controls yeast differentiation by interacting with kelch repeat proteins that mimic Gβ subunits', Molecular Cell, 10(1), pp. 163–173. doi: 10.1016/S1097-2765(02)00569-5.

Hong, K.-K., Vongsangnak, W., Vemuri, G. N. and Nielsen, J. (2011) 'Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis', Proceedings of the National Academy of Sciences, 108(29), pp. 12179–12184. doi: 10.1073/pnas.1103219108.

Hong, K. K. and Nielsen, J. (2013) 'Adaptively evolved yeast mutants on galactose show trade-offs in carbon utilization on glucose', Metabolic Engineering. Elsevier, 16(1), pp. 78–86. doi: 10.1016/j.ymben.2013.01.007.

Kvitek, D. J. and Sherlock, G. (2013) 'Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment.', PLoS genetics. Public Library of Science, 9(11), p. e1003972. doi: 10.1371/journal.pgen.1003972.

Li, Y., Venkataram, S., Agarwala, A., Dunn, B., Petrov, D. A., Sherlock, G. and Fisher, D. S. (2018) 'Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions', Current Biology. Elsevier Ltd., 28(4), p. 515–525.e6. doi: 10.1016/j.cub.2018.01.009.

Long, A., Liti, G., Luptak, A. and Tenaillon, O. (2015) 'Elucidating the molecular architecture of adaptation via evolve and resequence experiments', Nature Reviews Genetics. Nature Publishing Group, 16(10), pp. 567–582. doi: 10.1038/nrg3937.

Martínez, J. L., Bordel, S., Hong, K. K. and Nielsen, J. (2014) 'Gcn4p and the Crabtree effect of yeast: Drawing the causal model of the Crabtree effect in Saccharomyces cerevisiae and explaining evolutionary trade-offs of adaptation to galactose through systems biology', FEMS Yeast Research, 14(4), pp. 654–662. doi: 10.1111/1567-1364.12153.

Peeters, T., Louwet, W., Gelade, R., Nauwelaers, D., Thevelein, J. M. and Versele, M. (2006) 'Kelch-repeat proteins interacting with the G protein Gpa2 bypass adenylate cyclase for direct regulation of protein kinase A in yeast', Proceedings of the National Academy of Sciences, 103(35), pp. 13034–13039. doi: 10.1073/pnas.0509644103.

Schmitt, M. E., Brown, T. A. and Trumpower, B. L. (1990) 'A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae', Nucleic Acids Research, 18(10), pp. 3091–3092. doi: 10.1093/nar/18.10.3091.

Shirra, M. K., McCartney, R. R., Zhang, C., Shokat, K. M., Schmidt, M. C. and Arndt, K. M. (2008) 'A chemical genomics study identifies Snf1 as a repressor of GCN4 translation', Journal of Biological Chemistry, 283(51), pp. 35889–35898. doi: 10.1074/jbc.M805325200.

Tanaka, K., Nakafuku, M., Satoh, T., Marshall, M. S., Gibbs, J. B., Matsumoto, K., Kaziro, Y. and Toh-e, A. (1990) 'S. cerevisiae genes IRA1 and IRA2 encode proteins that may be functionally equivalent to mammalian ras GTPase activating protein', Cell, 60(5), pp. 803–807. doi: 10.1016/0092-8674(90)90094-U.

Venkataram, S., Dunn, B., Li, Y., Fisher, D. S., Sherlock, G., Petrov, D. A., Venkataram, S., Dunn, B., Li, Y., Agarwala, A., Chang, J., Ebel, E. R., Blundell, J. R., Levy, S. F., Fisher, D. S. and Sherlock, G. (2016) 'Development of a Comprehensive Genotype-to- Fitness Map of Adaptation-Driving Mutations in Resource Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast', Cell, pp. 1–12. doi: 10.1016/j.cell.2016.08.002.

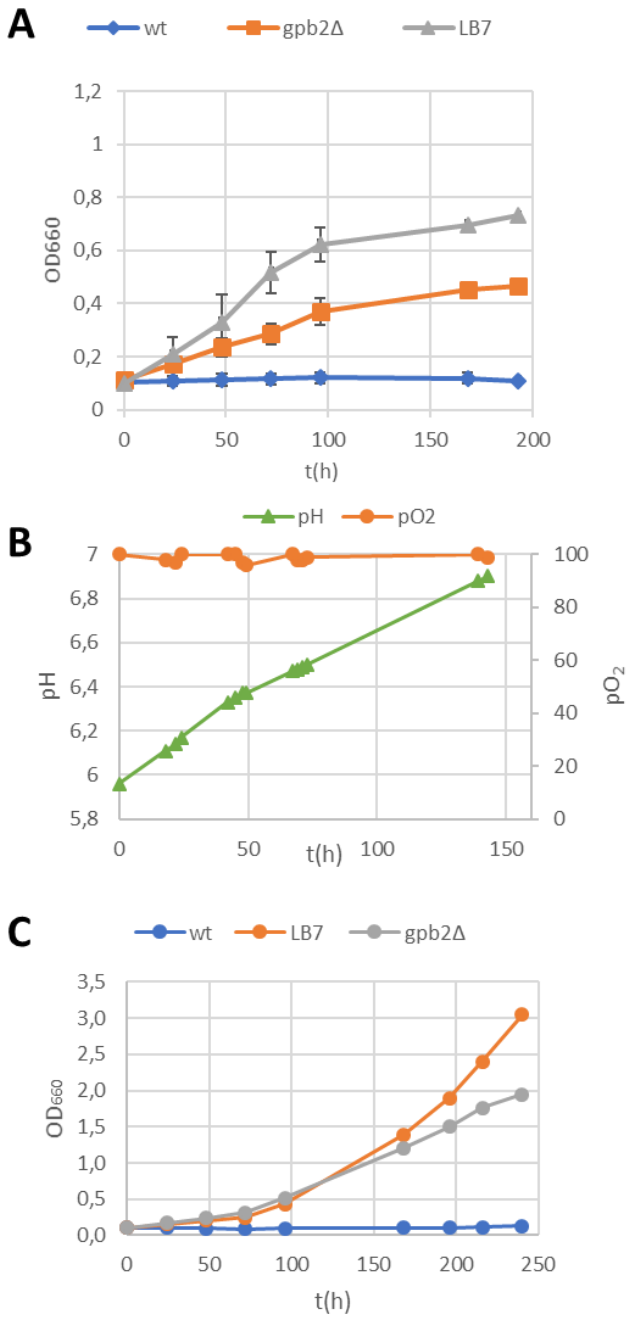
Verduyn, C., Postma, E., Scheffers, W. A. and Van Dijken, J. P. (1992) 'Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation', Yeast (Chichester, England), 8(7), pp. 501–517. doi: 10.1002/yea.320080703.

Wenger, J. W., Piotrowski, J., Nagarajan, S., Chiotti, K., Sherlock, G. and Rosenzweig, F. (2011) 'Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency.', PLoS genetics. Public Library of Science, 7(8), p. e1002202. doi: 10.1371/journal.pgen.1002202.

Zaman, S., Lippman, S. I., Schneper, L., Slonim, N. and Broach, J. R. (2009) 'Glucose regulates transcription in yeast through a network of signaling pathways.', Molecular systems biology. EMBO Press, 5(1), p. 245. doi: 10.1038/msb.2009.2.

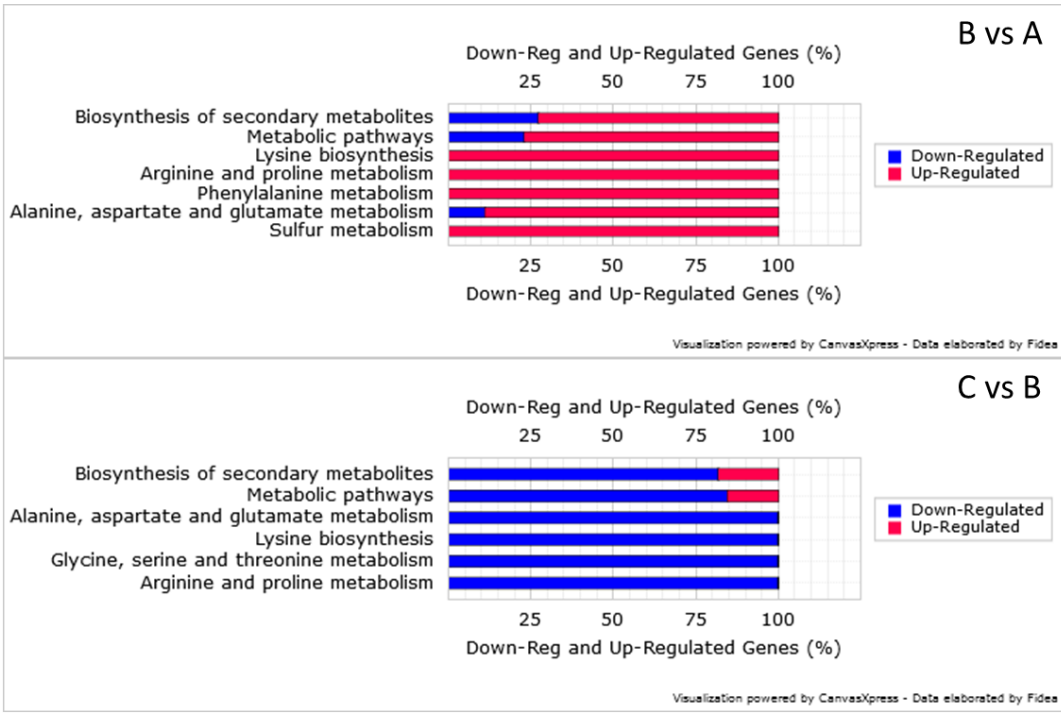
Zaman, S., Lippman, S. I., Zhao, X. and Broach, J. R. (2008) 'How Saccharomyces responds to nutrients.', Annual review of genetics, 42, pp. 27–81. doi: 10.1146/annurev.genet.41.110306.130206.

Supplementary information

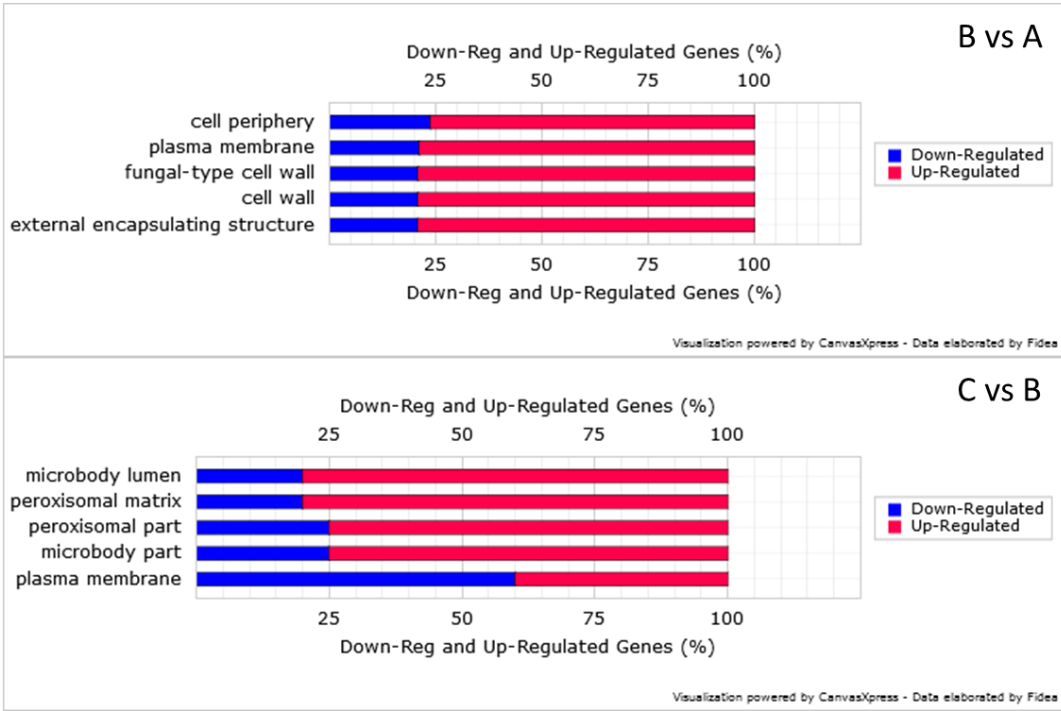


Supplementary figure S1 – (A) Growth gpb2Δ on glutamate as carbon and nitrogen source, compared to wild-type and LB7 strains. (B) Measure of pH and oxygenation for growth in fermentor without control of those parameters. (D) Growth of wt, LB7 and gpb2Δ on glutamate only shaking flask, buffered at pH 5.5 with 100mM MES.

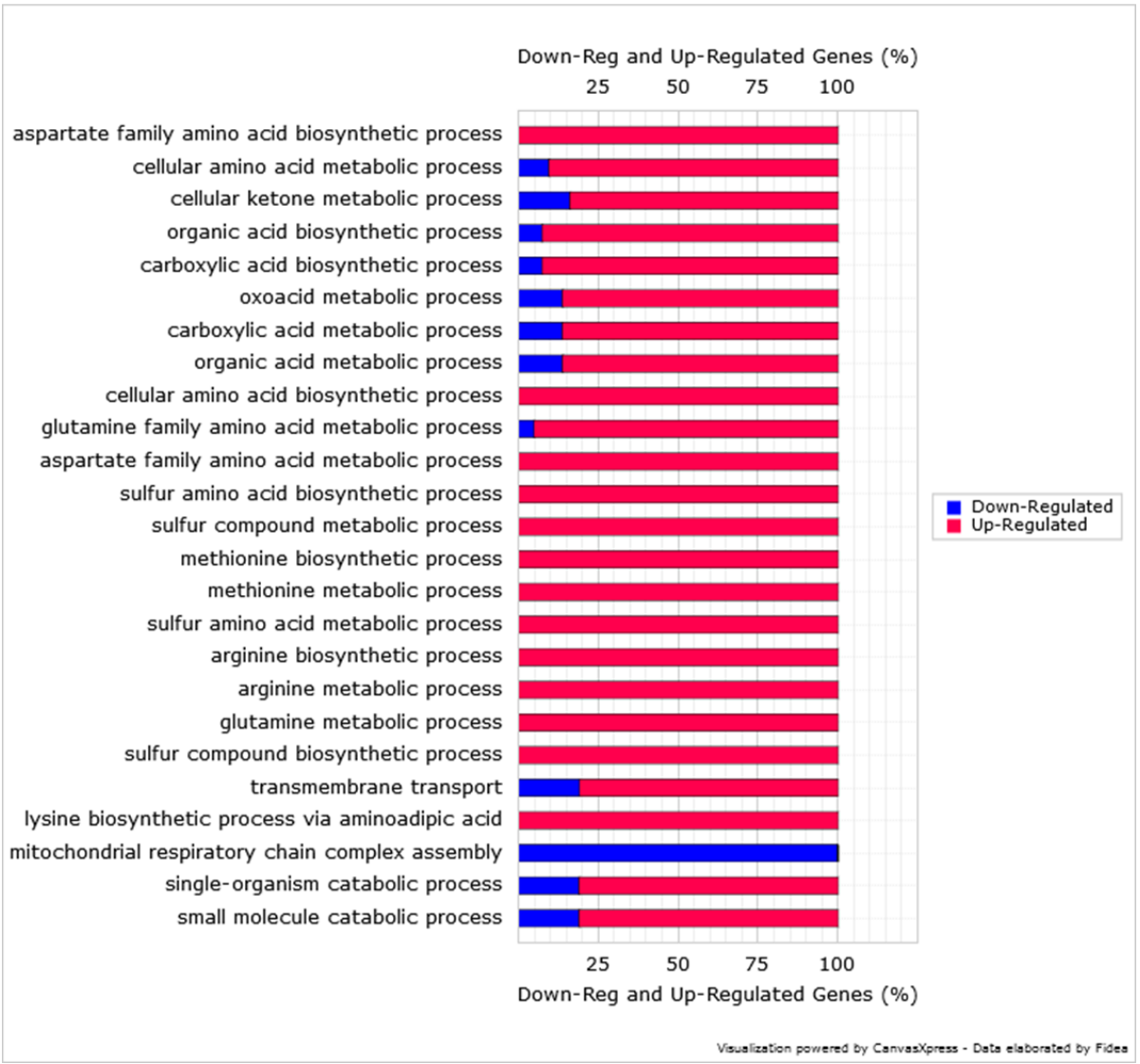
A FIDEA – KEGG



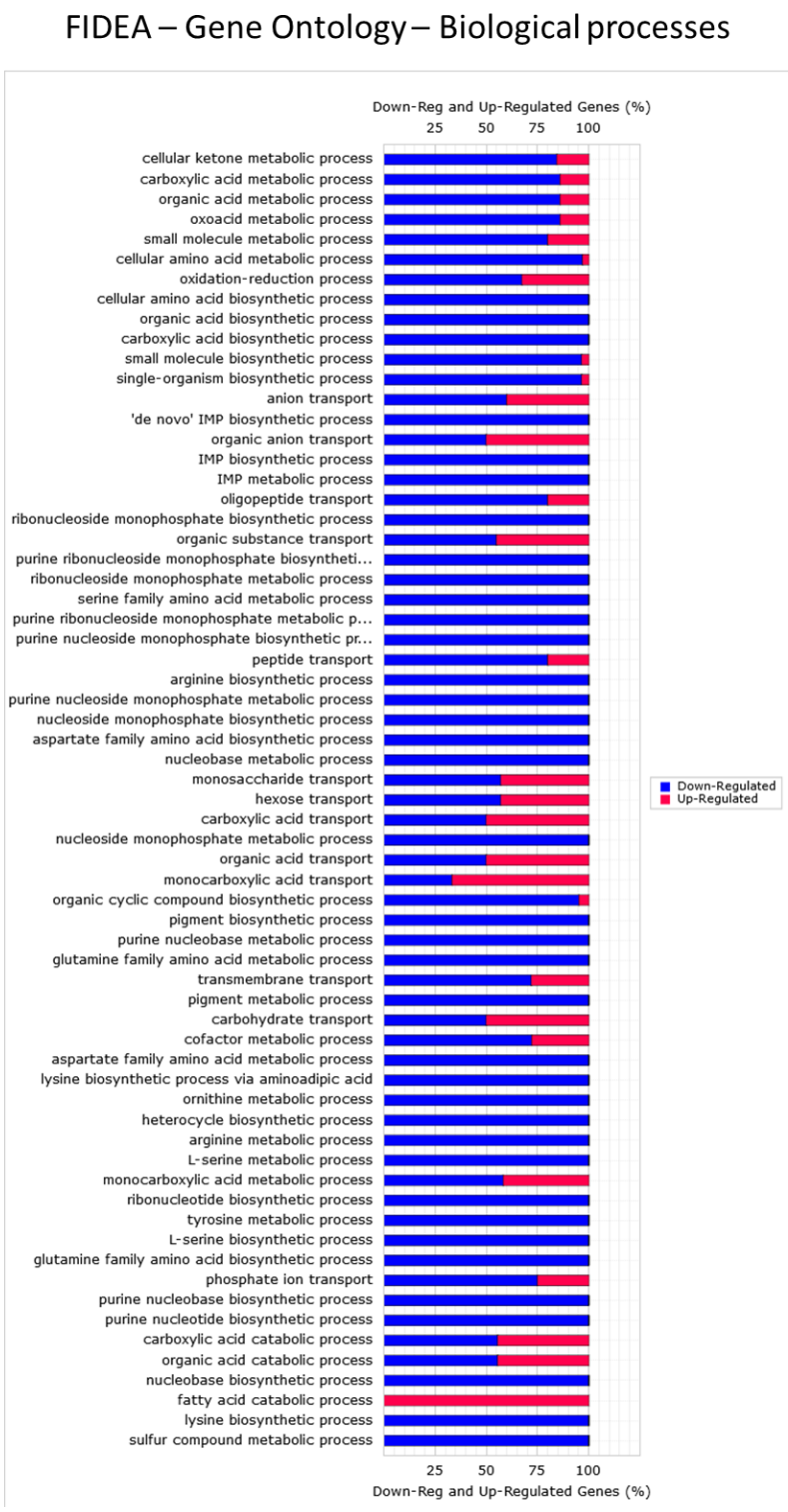
B FIDEA – Gene Ontology – Cellular compartments



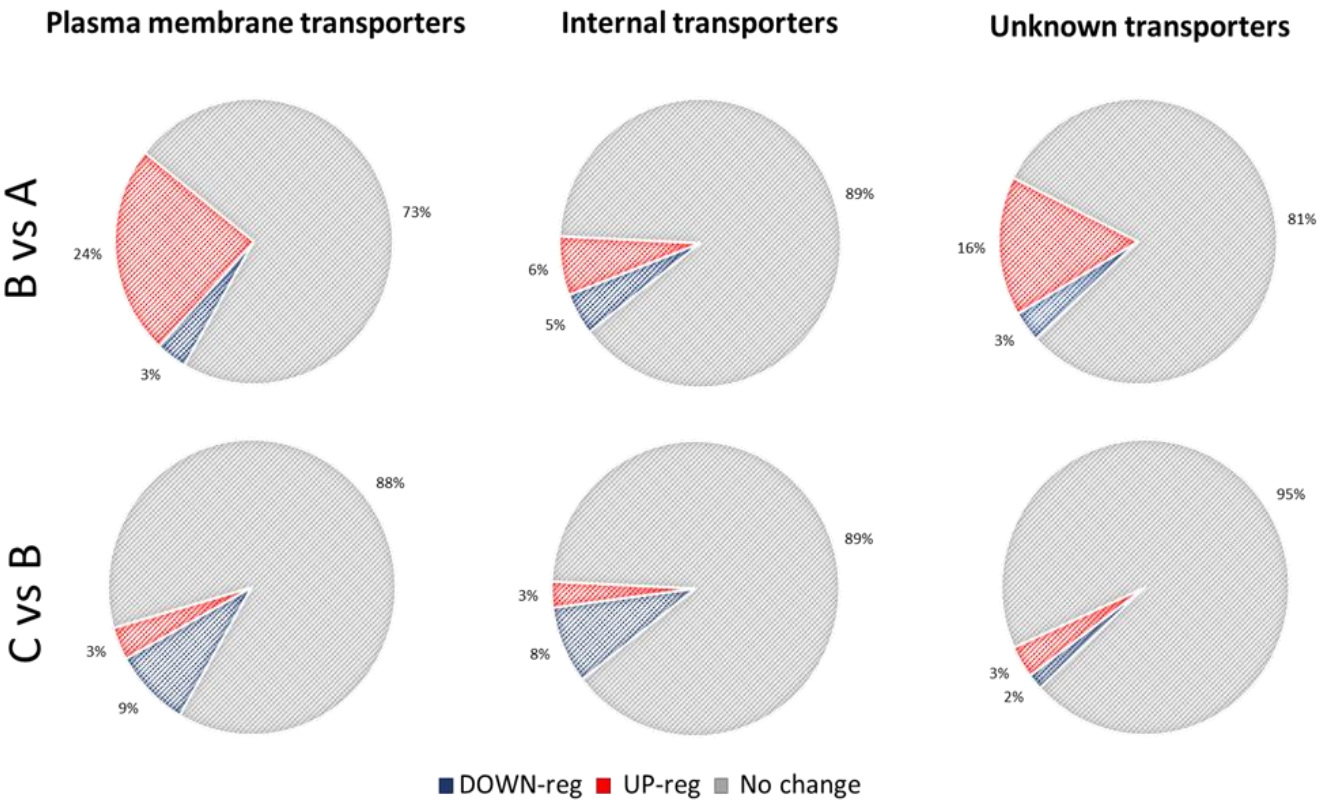
Supplementary figure S2 – KEGG (A) and GO cellular compartments (B) terms significantly associated with the B vs A (top) and C vs B (bottom) DEGs lists.



Supplementary figure S3 - GO biological processes terms significantly associated with the B vs A DEGs list.



Supplementary figure S4 - GO biological processes terms significantly associated with the B vs A DEGs list



Supplementary figure S5 – Representation of the percentage of plasma membrane, internal and uncharcterized transporters (according to the YTP database) regulated in the B vs A (top) and C vs B (bottom) comparisons.

Supplementary table S1 – Transcription factors significantly associated with the B vs A gene list

Transcription Factor	% in our set	% in Yeastract	p-value
Sok2p	54.24%	15.60%	0.0000E+00
Bas1p	67.33%	14.55%	0.0000E+00
Xbp1p	33.90%	15.65%	0.0000E+00
Gcn4p	66.26%	14.36%	0.0000E+00
Met32p	17.26%	20.36%	0.0000E+00
Ace2p	83.20%	11.30%	0.0000E+00
Cin5p	46.53%	13.85%	0.0000E+00
Swi5p	41.91%	13.98%	0.0000E+00
Mig3p	37.90%	14.73%	0.0000E+00
Zap1p	36.98%	15.59%	0.0000E+00
Ume6p	38.06%	16.78%	0.0000E+00
Msn2p	64.71%	12.29%	0.0000E+00
Msn4p	52.54%	13.28%	0.0000E+00
Pdr1p	33.59%	15.14%	0.0000E+00
Tec1p	70.88%	12.05%	0.0000E+00
Ste12p	71.65%	12.07%	0.0000E+00
Yox1p	33.90%	15.40%	0.0000E+00
Gis1p	17.57%	29.08%	0.0000E+00
Aft1p	34.21%	15.79%	0.0000E+00
Yhp1p	31.59%	15.92%	0.0000E+00
Gat4p	9.24%	27.52%	0.0000E+00
Mss1p	0.31%	100.00%	0.0000E+00

Sko1p	25.42%	17.82%	0.0000E+00
Rlm1p	21.88%	19.92%	0.0000E+00
Mit1p	0.15%	100.00%	0.0000E+00
Mfg1p	0.15%	100.00%	0.0000E+00
Com2p	0.15%	100.00%	0.0000E+00
Dig2p	0.15%	100.00%	0.0000E+00
Sfp1p	74.88%	11.21%	1.0000E-15
Pho2p	26.35%	15.91%	1.0000E-15
Yap6p	24.65%	16.39%	1.0000E-15
Flo8p	24.50%	16.26%	3.0000E-15
Yap1p	62.87%	11.81%	4.0000E-15
Rph1p	19.26%	17.58%	1.3000E-14
Arr1p	34.82%	14.11%	1.4000E-14
Mot3p	16.64%	18.34%	6.0000E-14
Stp2p	17.57%	17.70%	1.4200E-13
Pdr3p	26.96%	14.86%	4.9000E-13
Mig1p	13.56%	19.43%	5.3300E-13
Upc2p	10.63%	21.70%	8.0300E-13
Nrg1p	18.64%	16.71%	1.6240E-12
Stp1p	18.49%	16.62%	3.0330E-12
Rox1p	21.73%	15.61%	3.9870E-12
Adr1p	20.34%	15.94%	4.8890E-12
Crz1p	16.64%	16.90%	1.4878E-11
Cup2p	15.25%	17.10%	5.7994E-11
Eds1p	5.70%	27.21%	1.3553E-10

Met4p	31.74%	13.30%	1.6292E-10
Met31p	11.56%	18.38%	4.6289E-10
Hsf1p	34.21%	12.77%	1.2591E-09
Sut1p	14.02%	16.55%	2.1758E-09
Mga1p	16.80%	15.48%	2.7946E-09
Met28p	7.55%	21.30%	2.8845E-09
Mig2p	9.09%	19.22%	5.6748E-09
Rme1p	11.25%	17.51%	7.5111E-09
Swi4p	22.80%	13.83%	1.0913E-08
Opi1p	10.48%	17.80%	1.2027E-08
Leu3p	15.72%	15.32%	1.7379E-08
Cbf1p	30.05%	12.79%	1.9630E-08
Hap4p	21.42%	13.93%	2.1873E-08
Ixr1p	33.28%	12.46%	2.2000E-08
Rpi1p	6.01%	22.41%	2.2630E-08
Rfx1p	16.33%	15.06%	2.2866E-08
Gcr2p	26.50%	13.09%	3.4867E-08
Kar4p	18.95%	14.29%	3.7680E-08
Gzf3p	8.32%	18.69%	6.6253E-08
Ash1p	54.39%	11.07%	8.1872E-08
Oaf1p	19.11%	13.90%	1.6517E-07
Skn7p	21.88%	13.42%	1.6666E-07
Rap1p	51.46%	11.13%	1.8285E-07
Spt23p	34.05%	12.09%	2.0326E-07
Cad1p	17.41%	14.16%	2.3785E-07

Fhl1p	24.50%	12.74%	8.4432E-07
Phd1p	16.49%	14.01%	9.0174E-07
Ino4p	27.73%	12.30%	1.5313E-06
Sfl1p	4.78%	20.95%	2.4824E-06
Rgm1p	9.71%	15.91%	2.6503E-06
Dal80p	5.86%	18.91%	3.5292E-06
Lys14p	0.77%	71.43%	3.5878E-06
Tos8p	9.86%	15.69%	3.6435E-06
Rtg3p	14.33%	13.78%	9.8742E-06
Tbs1p	3.54%	22.55%	9.9979E-06
Rtg1p	7.40%	16.55%	1.2325E-05
YPR015C	4.31%	20.29%	1.3172E-05
Stp3p	1.23%	42.11%	1.6201E-05
Stp4p	5.08%	18.75%	1.6474E-05
Ecm22p	11.71%	14.29%	1.7711E-05
Oaf3p	9.71%	15.00%	1.9136E-05
Gln3p	22.50%	12.29%	2.0839E-05
Ino2p	13.87%	13.62%	2.2196E-05
Pho4p	25.89%	11.92%	2.5970E-05
Abf1p	46.84%	10.80%	2.7099E-05
Hot1p	3.39%	21.57%	3.1819E-05
Aft2p	11.40%	14.15%	3.1822E-05
Rpn4p	31.59%	11.47%	3.3193E-05
Wtm2p	4.78%	18.34%	4.4097E-05

Supplementary table S2 - Transcription factors significantly associated with the C vs B gene list

Transcription Factor	% in our set	% in Yeastract	p-value
Bas1p	73.66%	5.49%	0.0000E+00
Sok2p	62.05%	6.16%	0.0000E+00
Gcn4p	79.02%	5.91%	0.0000E+00
Mss1p	0.89%	100.00%	0.0000E+00
Gis1p	21.88%	12.50%	0.0000E+00
Mit1p	0.45%	100.00%	0.0000E+00
Mfg1p	0.45%	100.00%	0.0000E+00
Met4p	45.09%	6.52%	1.0000E-15
Zap1p	44.64%	6.50%	1.0000E-15
Pdr1p	41.52%	6.46%	3.7000E-14
Msn4p	58.93%	5.14%	4.2500E-13
Arr1p	43.30%	6.05%	5.0300E-13
Aft1p	39.73%	6.33%	6.2900E-13
Ume6p	40.18%	6.11%	3.4870E-12
Adr1p	27.68%	7.49%	5.9520E-12
Sfp1p	81.25%	4.20%	6.5010E-12
Swi5p	47.77%	5.50%	9.4120E-12
Rox1p	29.02%	7.20%	9.8070E-12
Cin5p	51.34%	5.27%	1.4815E-11
Pho2p	32.14%	6.70%	1.8496E-11
Ace2p	85.71%	4.02%	2.4140E-11
Msn2p	68.75%	4.51%	4.8651E-11
Yap1p	69.20%	4.48%	5.7126E-11
Pdr3p	33.48%	6.37%	7.5623E-11

Rlm1p	24.11%	7.57%	1.1955E-10
Hsf1p	42.41%	5.46%	5.3622E-10
Gat4p	11.61%	11.93%	6.8235E-10
Sko1p	27.68%	6.70%	7.3891E-10
Mig1p	17.41%	8.61%	1.4571E-09
Cad1p	24.55%	6.89%	2.8951E-09
Rpn4p	41.96%	5.26%	5.9098E-09
Yap6p	27.68%	6.35%	6.3019E-09
Hot1p	7.14%	15.69%	1.3757E-08
Kar4p	25.00%	6.50%	1.7402E-08
Yox1p	35.27%	5.53%	2.0718E-08
Gzf3p	12.50%	9.69%	2.1158E-08
Met32p	18.30%	7.45%	4.0924E-08
Tec1p	70.54%	4.14%	4.5223E-08
Cbf1p	36.16%	5.31%	8.1953E-08
Ste12p	70.54%	4.10%	9.8472E-08
Flo8p	26.34%	6.03%	1.0976E-07
Stp2p	19.64%	6.83%	1.7480E-07
Yap5p	24.11%	6.16%	2.0764E-07
Rme1p	14.73%	7.91%	2.0994E-07
Opi1p	13.84%	8.12%	2.7134E-07
Mot3p	18.30%	6.96%	2.8219E-07
Stp1p	20.98%	6.51%	2.8339E-07
Nrg1p	20.98%	6.49%	3.0810E-07
Sut1p	17.41%	7.09%	3.4088E-07

Yhp1p	31.25%	5.43%	3.7475E-07
Crz1p	19.20%	6.73%	3.7510E-07
Met31p	14.29%	7.84%	3.9110E-07
Xbp1p	33.04%	5.26%	5.7257E-07
Mig3p	36.61%	4.91%	2.2027E-06
Oaf3p	13.84%	7.38%	2.2379E-06
Cat8p	9.38%	9.17%	2.5279E-06
Met28p	9.38%	9.13%	2.7184E-06
Rph1p	19.64%	6.19%	2.8538E-06
Nrg2p	12.05%	7.78%	3.4232E-06
Stb5p	27.68%	5.33%	4.1979E-06
Eds1p	6.70%	11.03%	4.8161E-06
Upc2p	11.16%	7.86%	6.2091E-06
Dal81p	12.05%	7.32%	1.1116E-05
Fhl1p	28.12%	5.05%	2.1414E-05
Hap4p	23.66%	5.31%	2.7011E-05
Mig2p	10.27%	7.49%	3.0475E-05
Hap1p	11.16%	7.18%	3.0761E-05
Lys14p	1.34%	42.86%	3.0768E-05

Conclusions

In the two chapters of this thesis we have analysed in depth the different ways in which glutamate can be used by wild-type and mutant cells for different cellular purposes. At their core, both chapters place glutamate carbon backbone at the centre of its specificity. Glutamate is just an enzymatic reaction away from ammonium, the standard nitrogen source used in synthetic yeast media and alpha-ketoglutarate, a key intermediate of TCA cycle (Magasanik and Kaiser, 2002). AKG also results from the many transamination reactions in which glutamate is the donor molecule (Cooper, 1982).

The position of glutamate in the cellular metabolism reflects in the many ways in which we have shown it modifies and interacts with the cellular machinery. For instance, the interaction with the different carbon source is not just a matter of respiratory versus fermentative metabolism and glucose repression but is due to a different allocation of metabolic fluxes. On the other hand, chapter 2 mutants present alterations in the PKA pathway, which responds primarily in sensing glucose (Busti *et al.*, 2010) and elicits many of the cellular responses associated with presence of the sugar (Zaman *et al.*, 2009). Nonetheless, transceptors-based sensing of nitrogen and other nutrients after starvation is reported to activate PKA when glucose and all the other nutrients required for fermentative growth are present (Steyfkens *et al.*, 2018).

We showed that *S. cerevisiae* is capable of using the carbon atoms of glutamate for accumulation of biomass. Wild-type budding yeast C-catabolic use of glutamate is confined to post-diauxic growth on ethanol, while during fermentation of glucose a great part of AKG is secreted in the growth medium. In addition, glutamate carbons are incorporated into storage molecules which sustain an enhanced stress resistance. On the other hands, activating mutations in the protein kinase A signalling pathway are required for growth on glutamate as both a carbon and nitrogen source. We showed that this growth is strongly dependent on the pH of the growth medium and mutant cells are characterized by up-regulation of genes involved in biosynthesis and transport of amino acids.

It should already be obvious, and it is further strengthen by the results presented in this thesis, that the traditional biochemical representation of distinct metabolic pathways for utilization and anabolism of carbon, nitrogen and other nutrients is far from what happens in the cell. Dealing with this notion can be confusing, and it wold be a struggle to grasp a complete picture of the cell with a classical cell biology approach. For this reason, throughout the two chapters, a lot of information came from the use of -omics techniques (especially transcriptomics) and constraint-based modelling.

It should be noted that this kind of enquiry is that none of this analysis is self-sustaining. Integration of high-throughput data, physiological measurements and computational simulations is crucial. Without genomic data we would not have been able to reconstruct the metabolic network, and without constraints derived from chemostat experiments we would not have had a validation for its accuracy. Finally, without validation from actual data computational simulations would not have had the same impact. On the other hand, the second chapter probably would not have existed at all, if it wasn't for the result of the computational simulations, which showed the theoretical capacity of *S. cerevisiae* to use glutamate as carbon, nitrogen and energy source.

Yeast is particularly well suited for flux balance analysis, because continuous chemostat cultures are as close as it is possible to get to the steady state which is the underlying assumption of FBA. However, metabolic models and computational simulations are widely used also for multicellular eukaryotes. The paper reported in the appendix of this thesis (Graudenzi *et al.*, 2018) takes many of the concepts here discussed and applies them to human cells. In this work, we manually reconstructed a model of human metabolism and we developed a data integration framework, for the integration of transcriptomic data. Starting from these kind of data, scores can be assigned to the different metabolic reactions present in the model, which were used to cluster cancer patients according to their metabolic status and predict their prognosis.

This approach is similar to that seen in chapter 1 for the analysis of the reporter metabolites. This shows the flexibility of metabolic models, which do not necessarily require to compute the simulations to provide useful information about biological systems.

In conclusion, our work with glutamate not only provided answers and interpretations of yeast metabolic and signalling strategies but offers an example of combination of different techniques, which is able to deliver a system-level interpretation of biological data.

References

Atkinson, D. E. and Walton, G. M. (1967) 'Adenosine triphosphate conservation in metabolic regulation rat liver citrate cleavage enzyme', Journal of Biological Chemistry. ASBMB, 242(13), pp. 3239–3241.

Avendano, A. et al. (1997) 'GDH3 encodes a glutamate dehydrogenase isozyme, a previously unrecognized route for glutamate biosynthesis in Saccharomyces cerevisiae', Journal of Bacteriology, 179(17), pp. 5594–5597. doi: 10.1128/jb.179.17.5594-5597.1997.

Bakker, B. M. et al. (2001) 'Stoichiometry and compartmentation of NADH metabolism in Saccharomyces cerevisiae', FEMS Microbiology Reviews, 25(1), pp. 15–37. Available at: http://dx.doi.org/10.1111/j.1574-6976.2001.tb00570.x.

Bianconi, M. (2003) 'Calorimetric determination of thermodynamic parameters of reaction reveals different enthalpic compensations of the yeast hexokinase isozymes.', The Journal of biological chemistry, 278(21), pp. 18709–18713.

Van Der Bliek, A. M., Sedensky, M. M. and Morgan, P. G. (2017) 'Cell biology of the mitochondrion', Genetics, 207(3), pp. 843–871. doi: 10.1534/genetics.117.300262.

Bos, J. L. (1989) 'ras Oncogenes in Human Cancer : A Review ras Oncogenes in Human Cancer : A Review', Cancer Research, 49, pp. 4682–4689.

Burgard, A. P. and Maranas, C. D. (2003) 'Optimization-based framework for inferring and testing hypothesized metabolic objective functions', Biotechnology and Bioengineering, 82(6), pp. 670–677. doi: doi:10.1002/bit.10617.

Busti, S. et al. (2010) 'Glucose signaling-mediated coordination of cell growth and cell cycle in Saccharomyces cerevisiae.', Sensors (Basel, Switzerland). Molecular Diversity Preservation International, 10(6), pp. 6195–6240. doi: 10.3390/s100606195.

Campero-Basaldua, C. et al. (2017) 'Diversification of the kinetic properties of yeast NADP-glutamate-dehydrogenase isozymes proceeds independently of their evolutionary origin', MicrobiologyOpen, 6(2), pp. 1–18. doi: 10.1002/mbo3.419.

Cazzaniga, P. et al. (2014) 'Computational strategies for a system-level understanding of metabolism', Metabolites. Multidisciplinary Digital Publishing Institute, 4(4), pp. 1034–1087. doi: 10.3390/metabo4041034.

Compagno, C., Dashko, S. and Piškur, J. (2014) 'Introduction to Carbon Metabolism in Yeast', in Piškur, J. and Compagno, C. (eds) Molecular Mechanisms in Yeast Carbon Metabolism. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–19. doi: 10.1007/978-3-642-55013-3_1.

Conrad, M. et al. (2014) 'Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae.', FEMS microbiology reviews. The Oxford University Press, 38(2), pp. 254–99. doi: 10.1111/1574-6976.12065.

Cooper, T. G. (1982) Nitrogen Metabolism in Saccharomyces cerevisiae., The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. doi: 10.1101/087969180.11B.39.

DeLuna, A. et al. (2001) 'NADP-glutamate dehydrogenase isoenzymes of Saccharomyces cerevisiae: Purification, kinetic properties, and physiological roles', Journal of Biological Chemistry, 276(47), pp. 43775–43783. doi: 10.1074/jbc.M107986200.

Van Dijck, P. et al. (2017) 'Nutrient Sensing at the Plasma Membrane of Fungal Cells.', Microbiology spectrum, 5(2), pp. 1–21. doi: 10.1128/microbiolspec.FUNK-0031-2016.

Eleutherio, E. et al. (2014) 'Revisiting yeast trehalose metabolism', Current Genetics. Springer Berlin Heidelberg, 61(3), pp. 263–274. doi: 10.1007/s00294-014-0450-1.

Eltschinger, S. and Loewith, R. (2016) 'TOR Complexes and the Maintenance of Cellular Homeostasis', Trends in Cell Biology, pp. 148–159. doi: 10.1016/j.tcb.2015.10.003.

Faubert, B. et al. (2015) 'The AMP-activated protein kinase (AMPK) and cancer: Many faces of a metabolic regulator', Cancer Letters. Elsevier Ireland Ltd, 356(2), pp. 165–170. doi: 10.1016/j.canlet.2014.01.018.

Feist, A. M. and Palsson, B. O. (2010) 'The biomass objective function', Current Opinion in Microbiology, 13(3), pp. 344–349. doi: 10.1016/j.mib.2010.03.003.

Fernández, E., Moreno, F. and Rodicio, R. (1992) 'The ICL1 gene from Saccharomyces cerevisiae', European Journal of Biochemistry. Wiley/Blackwell (10.1111), 204(3), pp. 983–990. doi: 10.1111/j.1432-1033.1992.tb16720.x.

François, J. M., Walther, T. and Parrou, J. L. (2012) 'Genetics and Regulation of Glycogen and Trehalose Metabolism in Saccharomyces cerevisiae', Microbial Stress Tolerance for Biofuels. Edited by Z. L. Liu. Berlin, Heidelberg: Springer Berlin Heidelberg, 22, pp. 29–56. doi: 10.1007/978-3-642-21467-7.

Francois, J. and Parrou, J. L. (2001) 'Reserve carbohydrates metabolism in the yeast Saccharomyces cerevisiae', FEMS Microbiology Reviews, 25, pp. 125–145.

Garre, E. and Matallana, E. (2009) 'The three trehalases Nth1p, Nth2p and Ath1p participate in the mobilization of intracellular trehalose required for recovery from

saline stress in Saccharomyces cerevisiae', Microbiology, 155(9), pp. 3092–3099. doi: 10.1099/mic.0.024992-0.

Gasch, A. P. et al. (2000) 'Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes', Molecular Biology of the Cell, 11(12), pp. 4241–4257. doi: 10.1091/mbc.11.12.4241.

Gauthier, S. et al. (2008) 'Co-regulation of yeast purine and phosphate pathways in response to adenylic nucleotide variations', Molecular Microbiology, 68(6), pp. 1583–1594. doi: 10.1111/j.1365-2958.2008.06261.x.

Godard, P. et al. (2007) 'Effect of 21 different nitrogen sources on global gene expression in the yeast Saccharomyces cerevisiae.', Molecular and cellular biology, 27(8), pp. 3065–3086. doi: 10.1128/MCB.01084-06.

Görner, W. et al. (1998) 'Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity', pp. 586–597.

Graudenzi, A. et al. (2018) 'Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power', Journal of Biomedical Informatics. Elsevier, 87(July), pp. 37–49. doi: S1532046418301849.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', Cell. Elsevier Inc., 144(5), pp. 646–74. doi: 10.1016/j.cell.2011.02.013.

Hartig, A. et al. (1992) 'Differentially regulated malate synthase genes participate in carbon and nitrogen metabolism of S.cerevisiae', Nucleic Acids Research, 20(21), pp. 5677–5686. doi: 10.1093/nar/20.21.5677.

JE, S., VM, M. and CE, M. (1989) 'Isolation and characterization of OLE1, a gene affecting fatty acid desaturation from Saccharomyces cerevisiae.', The Journal of biological chemistry, 264(28), pp. 16537–16544.

Jones, E. and Fink, G. R. (1982) 'Regulation of amino acid and nucleotide biosynthesis in yeast', in The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression, pp. 181–299.

Jules, M. et al. (2004) 'Two Distinct Pathways for Trehalose Assimilation in the Yeast Saccharomyces cerevisiae Two Distinct Pathways for Trehalose Assimilation in the Yeast Saccharomyces cerevisiae', Applied and Environmental Microbiology, 70(5), pp. 2771–2778. doi: 10.1128/AEM.70.5.2771.

Jules, M. et al. (2008) 'New insights into trehalose metabolism by Saccharomyces cerevisiae: NTH2 encodes a functional cytosolic trehalase, and deletion of TPS1

reveals Ath1p-dependent trehalose mobilization', Applied and Environmental Microbiology, 74(3), pp. 605–614. doi: 10.1128/AEM.00557-07.

Kayikci, Ö. and Nielsen, J. (2015) 'Glucose repression in Saccharomyces cerevisiae', FEMS Yeast Research, 15(6), pp. 1–8. doi: 10.1093/femsyr/fov068.

Klug, L. and Daum, G. (2014) 'Yeast lipid metabolism at a glance', FEMS Yeast Research, 14(3), pp. 369–388. doi: 10.1111/1567-1364.12141.

Kohlwein, S. D., Veenhuis, M. and van der Klei, I. J. (2013) 'Lipid droplets and peroxisomes: Key players in cellular lipid homeostasis or a matter of fat-store'em up or burn'em down', Genetics, 193(1), pp. 1–50. doi: 10.1534/genetics.112.143362.

Lee, P. et al. (2008) 'Yeast Yak1 kinase, a bridge between PKA and stress-responsive transcription factors, Hsf1 and Msn2/Msn4.', Molecular microbiology. Blackwell Publishing Ltd, 70(4), pp. 882–95. doi: 10.1111/j.1365-2958.2008.06450.x.

Lee, P. et al. (2013) 'Rim15-dependent activation of Hsf1 and Msn2/4 transcription factors by direct phosphorylation in Saccharomyces cerevisiae', FEBS Letters, 587(22), pp. 3648–3655. doi: 10.1016/j.febslet.2013.10.004.

Legiša, M. (2014) 'Similarities and Differences Between Cancer and Yeast Carbohydrate Metabolism - Molecular Mechanisms in Yeast Carbon Metabolism', in Piškur, J. and Compagno, C. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 121–140. doi: 10.1007/978-3-642-55013-3_6.

Liu, Z. and Butow, R. A. (2006a) 'Mitochondrial Retrograde Signaling', Annual Review of Genetics, 40(1), pp. 159–185. doi: 10.1146/annurev.genet.40.110405.090613.

Liu, Z. and Butow, R. A. (2006b) 'Mitochondrial Retrograde Signaling', Annual Review of Genetics, 40(1), pp. 159–185. doi: 10.1146/annurev.genet.40.110405.090613.

Ljungdahl, P. O. and Daignan-Fornier, B. (2012) 'Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae', Genetics, 190(3), pp. 885–929. doi: 10.1534/genetics.111.133306.

Loewith, R. (2010) TORC1 signaling in budding yeast. 1st edn, Enzymes. 1st edn. Elsevier Inc. doi: 10.1016/S1874-6047(10)27009-9.

Long, A. et al. (2015) 'Elucidating the molecular architecture of adaptation via evolve and resequence experiments', Nature Reviews Genetics. Nature Publishing Group, 16(10), pp. 567–582. doi: 10.1038/nrg3937.

Magasanik, B. and Kaiser, C. A. (2002) 'Nitrogen regulation in Saccharomyces cerevisiae', Gene, 290(1–2), pp. 1–18. doi: 10.1016/S0378-1119(02)00558-9.

Marion, R. M. et al. (2004) ‘Sfp1 is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression’, Proceedings of the National Academy of Sciences, 101(40), pp. 14315–14322. doi: 10.1073/pnas.0405353101.

Nicastro, R. et al. (2017) ‘The architecture of the Rag GTPase signaling network’, Biomolecules, 7(3), pp. 1–21. doi: 10.3390/biom7030048.

Pedruzzi, I. et al. (2000) ‘Saccharomyces cerevisiae Ras/cAMP pathway controls post-diauxic shift element-dependent transcription through the zinc finger protein Gis1.’, The EMBO journal. EMBO Press, 19(11), pp. 2569–79. doi: 10.1093/emboj/19.11.2569.

Piškur, J. et al. (2006) ‘How did Saccharomyces evolve to become a good brewer?’, Trends in Genetics, 22(4), pp. 183–186. doi: 10.1016/j.tig.2006.02.002.

Ploier, B., Daum, G. and Petrovič, U. (2014) ‘Molecular Mechanisms in Yeast Carbon Metabolism: Lipid Metabolism and Lipidomics’, in Piškur, J. and Compagno, C. (eds) Molecular Mechanisms in Yeast Carbon Metabolism. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 169–215. doi: 10.1007/978-3-642-55013-3_8.

Pronk, J. T., Yde Steensma, H. and Van Dijken, J. P. (1996) ‘Pyruvate Metabolism in Saccharomyces cerevisiae’, Yeast. Wiley-Blackwell, 12(16), pp. 1607–1633. doi: 10.1002/(SICI)1097-0061(199612)12:16<1607::AID-YEA70>3.0.CO;2-4.

Radulovic, M. et al. (2013) ‘The emergence of lipid droplets in yeast: Current status and experimental approaches’, Current Genetics, 59(4), pp. 231–242. doi: 10.1007/s00294-013-0407-9.

Reinders, A. et al. (1998) ‘Saccharomyces cerevisiae cAMP-dependent protein kinase controls entry into stationary phase through the Rim15p protein kinase’, Genes Dev, 12(18), pp. 2943–2955.

Rødkaer, S. V et al. (2014) ‘Glucose- and nitrogen sensing and regulatory mechanisms in Saccharomyces cerevisiae.’, FEMS yeast research. The Oxford University Press, 14(5), pp. 683–96. doi: 10.1111/1567-1364.12157.

Rössler, H. et al. (2003) ‘Functional differentiation and selective inactivation of multiple Saccharomyces cerevisiae genes involved in very-long-chain fatty acid synthesis.’, Molecular genetics and genomics : MGG, 269(2), pp. 290–298.

Rutter, J., Winge, D. R. and Schiffman, J. D. (2010) ‘Succinate dehydrogenase - Assembly, regulation and role in human disease’, Mitochondrion. Mitochondria Research Society, 10(4), pp. 393–401. doi: 10.1016/j.mito.2010.03.001.

Sampaio-Marques, B. et al. (2011) 'Yeast chronological lifespan and proteotoxic stress: is autophagy good or bad?', Biochemical Society Transactions, 39(5), pp. 1466–1470. doi: 10.1042/BST0391466.

Sánchez, B. J. et al. (2017) 'Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints', pp. 1–16. doi: 10.15252/msb.20167411.

Shashkova, S., Welkenhuysen, N. and Hohmann, S. (2015) 'Molecular communication: Crosstalk between the Snf1 and other signaling pathways', FEMS Yeast Research, 15(4), pp. 1–10. doi: 10.1093/femsyr/fov026.

Smets, B. et al. (2010) Life in the midst of scarcity: Adaptations to nutrient availability in Saccharomyces cerevisiae, Current Genetics. doi: 10.1007/s00294-009-0287-1.

de Smidt, O., du Preez, J. C. and Albertyn, J. (2008) 'The alcohol dehydrogenases of Saccharomyces cerevisiae : a comprehensive review', FEMS Yeast Research, 8(7), pp. 967–978. doi: 10.1111/j.1567-1364.2008.00387.x.

Stelling, J. (2004) 'Mathematical models in microbial systems biology', Current Opinion in Microbiology, 7(5), pp. 513–518. doi: 10.1016/j.mib.2004.08.004.

Steyfkens, F. et al. (2018) 'Multiple transceptors for macro- and micro-nutrients control diverse cellular properties through the PKA pathway in yeast: A paradigm for the rapidly expanding world of eukaryotic nutrient transceptors up to those in human cells', Frontiers in Pharmacology, 9(MAR), pp. 1–22. doi: 10.3389/fphar.2018.00191.

Strijbis, K. and Distel, B. (2010) 'Intracellular acetyl unit transport in fungal carbon metabolism', Eukaryotic Cell, 9(12), pp. 1809–1815. doi: 10.1128/EC.00172-10.

Talarek, N. et al. (2010) 'Initiation of the TORC1-Regulated G0 Program Requires Igo1/2, which License Specific mRNAs to Evade Degradation via the 5′-3′ mRNA Decay Pathway', Molecular Cell. Elsevier, 38(3), pp. 345–355. doi: 10.1016/j.molcel.2010.02.039.

Tehlivets, O., Scheuringer, K. and Kohlwein, S. D. (2007) 'Fatty acid synthesis and elongation in yeast', Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids, 1771(3), pp. 255–270. doi: 10.1016/j.bbalip.2006.07.004.

Thon, V. J. et al. (1992) 'Coordinate Regulation of Glycogen Metabolism in the Yeast Saccharomyces cerevisiae', The Journal of biological chemistry, 267(21), pp. 15224–15228.

Torija, M. J. et al. (2005) ‘Glycogen synthesis in the absence of glycogenin in the yeast Saccharomyces cerevisiae’, FEBS Letters, 579(18), pp. 3999–4004. doi: 10.1016/j.febslet.2005.06.007.

Urban, J. et al. (2007) ‘Sch9 Is a Major Target of TORC1 in Saccharomyces cerevisiae’, Molecular Cell, 26(5), pp. 663–674. doi: 10.1016/j.molcel.2007.04.020.

Usaite, R. et al. (2009) ‘Reconstruction of the yeast Snf1 kinase regulatory network reveals its role as a global energy regulator’, Molecular Systems Biology. Nature Publishing Group, 5(319), pp. 1–12. doi: 10.1038/msb.2009.67.

Wenger, J. W. et al. (2011) ‘Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency.’, PLoS genetics. Public Library of Science, 7(8), p. e1002202. doi: 10.1371/journal.pgen.1002202.

De Wever, V. et al. (2005) ‘A dual role for PP1 in shaping the Msn2-dependent transcriptional response to glucose starvation’, EMBO Journal, 24(23), pp. 4115–4123. doi: 10.1038/sj.emboj.7600871.

Young, E. T. et al. (2003) ‘Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8’, Journal of Biological Chemistry, 278(28), pp. 26146–26158. doi: 10.1074/jbc.M301981200.

Zaman, S. et al. (2009) ‘Glucose regulates transcription in yeast through a network of signaling pathways.’, Molecular systems biology. EMBO Press, 5(1), p. 245. doi: 10.1038/msb.2009.2.

Zhang, W. et al. (2018) ‘Regulation of Sensing, Transportation, and Catabolism of Nitrogen Sources in Saccharomyces cerevisiae.’, Microbiology and molecular biology reviews : MMBR, 82(1), pp. e00040-17. doi: 10.1128/MMBR.00040-17.

Appendix

Paper: Alex Graudenzi, Davide Maspero, Marzia Di Filippo, Marco Gnugnoli, Claudio Isella, Giancarlo Mauri, Enzo Medico, Marco Antoniotti, Chiara Damiani, Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power, Journal of Biomedical Informatics, Volume 87, 2018, Pages 37-49, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2018.09.010.

(http://www.sciencedirect.com/science/article/pii/S1532046418301849)

Journal of Biomedical Informatics 87 (2018) 37–49

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power

Alex Graudenzi[a,1], Davide Maspero[b,1], Marzia Di Filippo[b,c], Marco Gnugnoli[b,c], Claudio Isella[d,e], Giancarlo Mauri[a,c], Enzo Medico[d,e], Marco Antoniotti[a,f], Chiara Damiani[a,c,*]

[a] Department of Informatics, Systems and Communication, University of Milan-Bicocca, Milan, Italy
[b] Department of Biotechnology and Biosciences, University Milano-Bicocca, Milan, Italy
[c] SYSBIO Centre of Systems Biology, University Milano-Bicocca, Milan, Italy
[d] University of Torino, Department of Oncology, Candiolo, Torino, Italy
[e] Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Torino, Italy
[f] Milan Center for Neuroscience, University of Milan-Bicocca, Monza, Italy

ARTICLE INFO

Keywords:
Metabolic networks
RNA-seq data
Genome-wide models
Sample stratification
Cancer metabolism

ABSTRACT

Effective stratification of cancer patients on the basis of their molecular make-up is a key open challenge. Given the altered and heterogenous nature of cancer metabolism, we here propose to use the overall expression of central carbon metabolism as biomarker to characterize groups of patients with important characteristics, such as response to *ad-hoc* therapeutic strategies and survival expectancy.

To this end, we here introduce the data integration framework named *Metabolic Reaction Enrichment Analysis* (MaREA), which strives to characterize the metabolic deregulations that distinguish cancer phenotypes, by projecting RNA-seq data onto metabolic networks, without requiring metabolic measurements. MaREA computes a score for each network reaction, based on the expression of the set of genes encoding for the associated enzyme(s). The scores are first used as features for cluster analysis and then to rank and visualize in an organized fashion the metabolic deregulations that distinguish cancer sub-types.

We applied our method to recent lung and breast cancer RNA-seq datasets from The Cancer Genome Atlas and we were able to identify subgroups of patients with significant differences in survival expectancy. We show how the prognostic power of MaREA improves when an extracted and further curated core model focusing on central carbon metabolism is used rather than the genome-wide reference network.

The visualization of the metabolic differences between the groups with best and worst prognosis allowed to identify and analyze key metabolic properties related to cancer aggressiveness. Some of these properties are shared across different cancer (sub) types, e.g., the up-regulation of nucleic acid and amino acid synthesis, whereas some other appear to be tumor-specific, such as the up- or down-regulation of the phosphoenolpyruvate carboxykinase reaction, which display different patterns in distinct tumor (sub)types.

These results might be soon employed to deliver highly automated diagnostic and prognostic strategies for cancer patients.

1. Introduction

Alterations of energy metabolism play a relevant role in several pathologies, such as metabolic syndrome, ageing, cancer, diabetes and neurodegeneration [1]. Current research on human metabolism typically relies on genome-wide reconstructions of human metabolic networks [2], such as *Human Metabolic Reaction* (HMR) [3] and Recon [4,5]. These models include most metabolic reactions that may occur in a generic cell, as well as Gene-Protein-Reaction (GPR) associations, which are logical formulas that describe how gene products concur to catalyze a given reaction.

Several strategies have been proposed to integrate *-omics* data into metabolic networks by exploiting GPRs, in order to derive context-specific models, e.g., the active metabolic network in a given cell or tissue [6–8]. Such approaches are usually conceived in the framework of constraint-based modeling and, in particular, of *Flux Balance Analysis*

* Corresponding author at: Viale Sarca 336 I, 20126 Milan, MI, Italy.
E-mail address: chiara.damiani@unimib.it (C. Damiani).
[1] Equal contributors.

https://doi.org/10.1016/j.jbi.2018.09.010
Received 9 July 2018; Received in revised form 7 September 2018; Accepted 14 September 2018
Available online 20 September 2018
1532-0464/

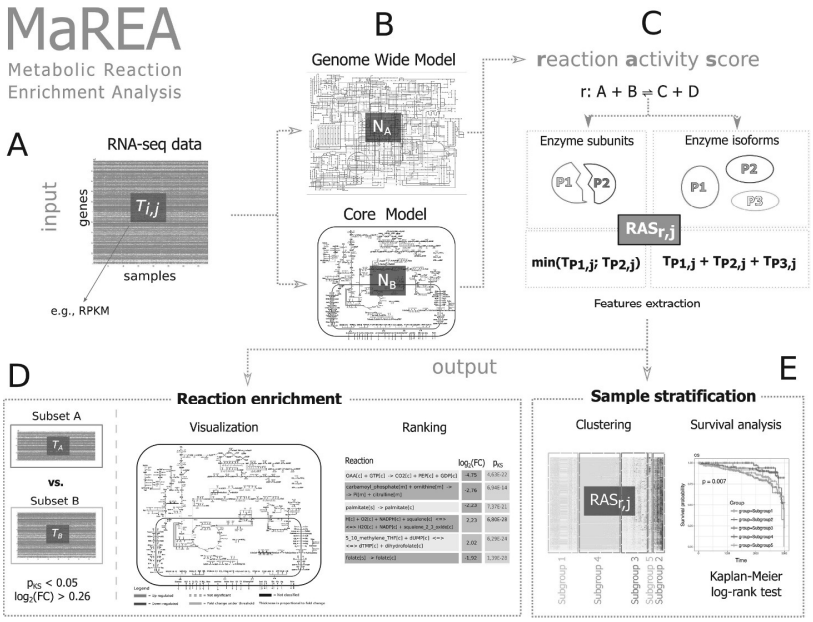


Fig. 1. MaREA pipeline. (A) MaREA takes as input a n sample × m genes matrix T which includes the normalized read count of each gene from a given cross-sectional RNA-seq dataset. (B) MaREA can use different input metabolic reaction networks, e.g., the genome-wide Model, or any subset of it. (C) A *Reaction Activity Score* (RAS) is defined for any reaction r in the input network and any sample, by distinguishing the case of reactions involving enzymes composed by different subunits (in this case the RAS is computed as the minimum of the transcript level of the genes encoding the subunits), and that of reaction catalyzed by different enzyme isoforms (in this case the RAS is computed as the sum of the transcript level of the genes encoding the isoforms). (D) Given two distinct subsets T_A and T_B of the original dataset, the RASs of a given reaction in the two cases are compared; if the p-value of the Kolmogorov-Smirnov test is significant (<0.05 - default) and the $\log_2$ fold-change is larger than 0.263 (default), that reaction will be marked in the final graph as up- or down-regulated. Accordingly, a reaction ranking can be provided. (E) MaREA can stratify patients by employing the RAS as feature on standard clustering methods, e.g., k-means. Survival analyses, such as log-rank test on Kaplan-Meier curves, can finally provide a prognostic validation of the clusters (notice that the image is intended for explanatory purpose only and does not reproduce any real case study).

(FBA) [9]. FBA relies on *linear programming* techniques to compute the flux through each reaction under a *steady state assumption*, and requires metabolic measurements to constrain nutrient exchange. Unfortunately, the simultaneous presence of metabolic measurements and distinct -omics data on the same patient is rarely available in public databases, such as the *The Cancer Genome Atlas* (TCGA) [10]. Besides, the same metabolic constraints hardly hold for all patients within a single dataset, and more so in highly heterogenous diseases, such as cancer.

Moreover, FBA poses many modeling challenges, e.g., the definition of an appropriate objective function and unfeasibility problems (see Section 2 for a more detailed discussion).

To address many of these issues, we here introduce a novel data-integration framework named *Metabolic Reaction Enrichment Analysis* (MaREA) (Fig. 1), which focuses on *transcriptional deregulation* of metabolic reactions, rather than on metabolic flux estimation. That is, MaREA processes transcriptional data, such as RNA-seq, without requiring metabolic measurements.

For each reaction of a given metabolic network, MaREA computes a *Reaction Activity Score* (RAS), which describes the extent of its activity in a given condition, as a function of the expression of the genes encoding for the *subunits* and/or the *isoforms* of the associated enzyme(s). The RAS provides a more refined information than the mere list of genes associated with a reaction, without requiring the setting of any arbitrary threshold, or to binarize data (i.e., gene present or absent), as required by other approaches, such as [11]. Analogous scores are have been employed to integrate continuous gene expression data in constraint-based simulations [6–8]. However, in MaREA the RAS is not used to define constraints or objective functions in FBA simulations. Instead, it is used as a static representation of the metabolic deregulation of a given sample, which can be directly used to compare different sample sets, e.g., different patient cohorts, or physiological vs. pathological condition.

Moreover, the features extracted by MaREA can be used to stratify samples in an unsupervised manner (*Metabolic Feature Extraction*). Such stratification might provide relevant prognostic indications, as shown in the case studies in Section 3.

In summary, MaREA can be used to:

(i) rank reactions according to their activity variation across different phenotypes and/or experimental conditions.
(ii) Enrich the map of human metabolic routes with the RAS variation, providing a clear and user-friendly visualization of how deregulated paths are interconnected.
(iii) Efficiently stratify samples according to their metabolic activity, hence providing a new unsupervised clustering tool, with testable clinical relevance, which can be assessed, e.g., via standard survival analyses.

In order to test our approach, we applied MaREA to the investigation of cancer metabolic heterogeneity. The heterogeneity of cancer genotypes and phenotypes hinders the identification of targets for effective treatments and is a major cause of tumor relapse [12]. Therefore, it is common practice to statistically compare the gene expression of patient cohorts, based on clinical observations and/or molecular features, in order to understand how the hallmarks of cancer can be (alternatively) achieved in terms of gene expression regulation. A particularly relevant hallmark for cancer treatment is the *metabolic*

reprogramming of cancer cells [13,14].

In particular, we applied MaREA to two distinct publicly available datasets in the TCGA database [10]: (i) the TCGA-BRCA dataset on breast cancer [15], and (ii) the TCGA-LUAD dataset on lung adenocarcinoma [16]. In the analyses, we employed both the genome wide-model Recon 2.2 [4] and a manually curated subset of it, corresponding to the model of central carbon metabolism (HMRCore), previously used in [17,18].

Because the TCGA-BRCA dataset includes both cancer and normal biopsies for a subset of the samples, we first used MaREA to show its effectiveness in identifying, rank and visualize the enriched metabolic reactions between normal and cancer samples. This allowed to reproduce well-known features of cancer deregulation and produce new experimental hypotheses. Finally, we used MaREA to stratify cancer patients in distinct metabolic clusters with respect to both the TCGA-BRCA and the TCGA-LUAD datasets. Standard survival analysis highlighted statistically significant prognostic predictions for the identified clusters.

MaREA is freely available as a user-friendly MATLAB tool, which allows to process input transcriptomic data, e.g., RNA-seq, in various file formats, and input metabolic networks in COBRA-compliant file format [19], e.g., SBML (see "Availability" Section at the end of the paper).

2. Materials and methods

2.1. Input

MaREA takes as input any RNA-seq dataset in the form of a $n \times m$ matrix T, where n is the number of genes and m is the number of samples of the considered cohort (see Fig. 1-A). Each element $T_{i,j}$, $i = 1, ..., n$, $j = 1, ..., m$ corresponds to the *normalized read count* of gene i in sample j such as, for instance, the *RPKM* (Reads per Kilobase per Million mapped reads).

MaREA then filters T according to a specific input reaction network N, e.g., the genome-wide metabolic network Recon 2.2 [4] or any possible subset of it. In particular, we define the set of reactions as $\mathcal{R} = \{r \in N\}$. Therefore, T is filtered by retaining only the rows corresponding to genes that are associated with enzymes involved in the reactions included in $\mathcal{R}$ (see Fig. 1-B).

GPR logical formulas include *AND* and *OR* logical operators. AND rules are employed when distinct genes encode different *subunits* of the same enzyme, i.e., *all* the subunits are *necessary* for the reaction to occur. OR rules describe the scenario in which distinct genes encode *isoforms* of the same enzyme, i.e., either isoform is *sufficient* to catalyze the reaction.

For example, the succinate-Coenzyme A ligase enzyme is formed by the subunits alpha (gene SUCLG1) and beta gene (SUCLG2) and catalyzes the reaction $Pi + succinyl\text{-}CoA + GDP \leftrightarrow CoA + succinate + GTP$. The gene-enzyme rule for this reaction is therefore: SUCLG1 *AND* SUCLG2. Conversely, ACACA and ACACB are respectively fully functional enzyme for the reaction *acetyl-coenzyme A ligase carboxylase*, thus the rule is ACACA *OR* ACACB. Such logical operators can of course be combined to depict multi-protein catalytic complexes or more complex situations involving both subunits and isoforms. For instance, ribonucleotide reductase is formed by two subunits: the catalytic (M1) and the regulatory one. The latter exists in two isoforms (M2 and M2B). The rule for this enzyme will therefore be RRM1 *AND* (RRM2 *OR* RRM2B).

2.2. Reaction Activity Score (RAS)

To avoid the definition of arbitrary thresholds on the transcript level, we do not resolve the logical expressions in a Boolean fashion, but we employ a *Reaction Activity Score* (RAS), for each sample $j = 1, ..., m$, and each reaction $r \in \mathcal{R}$ (see Fig. 1-C). In particular, in order to compute the RAS we distinguish:

- Reactions with AND operator (i.e., enzyme subunits).

$$RAS_{r,j} = \min(T_{i,j}: i \in \mathcal{A}_r), \quad (1)$$

  where A_r is the set of genes that encode the subunits of the enzyme catalyzing reaction r.
- Reactions with OR operator (i.e., enzyme isoforms).

$$RAS_{r,j} = \sum_{i \in O_r} T_{i,j}, \quad (2)$$

  where O_r is the set of genes that encode isoforms of the enzyme that catalyzes reaction r.

In case of composite reactions, we respect the standard precedence of the two operators. Let $|\mathcal{R}|$ be the cardinality of the set of reactions, the final output is therefore a $|\mathcal{R}| \times m$ matrix M, where each element $M_{r,j}$ is the RAS computed for reaction r in sample j.

Similarly to what has been proposed in [20] to improve FBA predictions, the intuition underneath the introduction of the RAS is that enzyme isoforms contribute *additively* to the overall activity of a given reaction, whereas enzyme subunits *limit* its activity, by requiring all the components to be present for the reaction to occur.

Clearly, we are here adopting a deeply simplified approach to reaction network modeling, by neglecting, for instance, the great heterogeneity of reaction kinetic constants and protein binding affinities, of translation rates, and any possible post-transcriptional regulation effect that might occur within a cell. An optimal choice would be to weigh all the reactions according to such quantities, yet direct measurements or robust estimates are very rarely available, especially for genome-wide models. Therefore, at first approximation, we here assume that all enzyme isoforms and subunits contribute uniformly to the reaction activity of a given reaction, as we expect that this choice does not affect the up-/down-regulation interplay observed at the network level.

2.3. Reaction enrichment: visualization and ranking

One important feature of MaREA is the ability to identify and visualize in an explicit way the metabolic routes that are up- or down-regulated in different sample sets and/or experimental conditions (see Fig. 1-D).

Given two distinct RNA-seq datasets, or two partitions of the same dataset, T_A and T_B, and an input metabolic reaction network N, we first compute the RAS matrices M_A and M_B. For each reaction $r \in N$ we then perform a non-parametric two-sample *Kolmogorov-Smirnov* (KS) test with a default p-value threshold equal to 0.05, to verify whether the distributions of RASs over the samples in the two sets are significantly different.

In that case, we compute the log_2 fold-change of the average RAS_r in the two groups. Because KS-test considers as significantly different distributions with the same mean, but different standard deviation, by default we consider as relevant only log_2 fold-change larger than $\log_2(1.2) = 0.263$ (i.e., corresponding to a 20% variation of the average RAS). In line with the philosophy of GSEA [21], we use a relaxed threshold for the fold-change because even a difference of 20% in genes encoding members of a metabolic pathway may dramatically alter the flux through the pathway. The significance threshold on the p-value of the Kolmogorov-Smirnov test, on the other hand, ensures that expression distributions of the two groups are indeed different. Notice that MaREA allows the user to choose different values for the RAS p-value and the fold-change threshold.

Next, MaREA uses the significant RAS fold-changes to: (i) determine a ranking of the most relevant up- and down-regulated reactions in the two sets, (ii) map such quantities over the input metabolic network N, by respectively coloring in red/blue the up-/down-regulated reactions, and by setting the edge thickness as proportional to the RAS fold-change. The reactions that will either display non-significant p-value

(either due to identical distributions or to statistically insufficient sample size) or a RAS fold-change below the threshold will not be included in the ranking and will be marked in gray color on the metabolic network.

2.4. RAS-based sample stratification

Another advantage of our approach is that it is possible to employ the RAS as an effective feature to identify sample subgroups (or *clusters*) that share similar metabolic properties (see Fig. 1-E). In particular, MaREA includes a *k*-means clustering [22], which uses the RASs of all reactions $r \in R$ (normalized on patients) to identify sample clusters with distinct metabolic behaviors. Clusters can be compared by means of the reaction enrichment procedure described above, by ranking the significantly different reactions in the distinct clusters and visualizing the RAS fold-changes on the input network.

When data is available, clusters can also be tested via standard *survival analysis*, such as the *log-rank* test on *Kaplan-Meier* curves, hence providing an orthogonal validation of the clustering results with clinical relevance. We show how clustering on RASs indeed can produce significant prognostic predictions.

2.5. Metabolic network

To compute the RAS of cancer samples in a given TCGA dataset at the genome-wide scale, we used the GPRs included in the most up-to-date genome-wide network of human metabolism: Recon 2.2 [4]. In particular, in order to visualize MaREA results at the genome-wide level, we modified the graphical attributes of the model map in *xml* format obtained from the *Virtual Metabolic Human* (VMH) – https://vmh.uni.lu – which is readable by the tool *Cell Designer* [23].

To focus, instead, on central carbon metabolism, we used the metabolic core model (HMRcore) introduced in [18]. For the sake of completeness, we included in the model mitochondrial palmitate degradation and gluconeogenesis. As the original version of the model does not include information on GPRs, such rules have been extracted and manually curated from Recon 2.2 [4] and included in the HMRcore model. In particular, we verified the correctness of GPR rules taking into account the information retrieved from the *Human Protein Atlas* [24] for the protein tissue location, from *UniProtKB* [25] for the enzyme complex composition, and from *KEGG* [26] to check gene/enzyme association. We checked for possible inconsistencies within Recon 2.2 and 29 GPRs were corrected. All the corrections made on Recon 2.2 GPR associations are reported in Supplementary Table S1. Notably, we corrected for the rules for Complex I-V of the respiratory chain, which were too strict in their original formulation. In particular, most patients in the BRCA dataset would have a null RAS for Complex IV, as gene COX7B2 is not expressed in most patients. It is important to notice that if we perform a FBA simulation with a null upper bound for Complex IV reaction, we do not obtain any optimal solution with the HMRcore model. The original rule considers indeed many genes that are actually isoforms and not subunits, thus requiring an OR and not an AND operator. In particular, in the original formulation COX7B does not allow to substitute for COX7B2.

The final version of the HMRcore model includes 264 reactions with a GPR rules and 405 metabolic genes that are associate to them. Genes are identified with the *HGNC ID* provided by the *HUGO Gene Nomenclature Committee* [27]. The SBML of the model is provided in Supplementary file S1. The tabular description of the model is provided in Supplementary file S2.

It should be noted that not every reaction in the metabolic models is associated with a gene-enzyme rule: for instance some reactions have been included to fill the gaps in steady state computations, but we lack knowledge on the associate genes. In detail, 4742 (263) reactions over 7785 (314) are associated with a gene-enzyme rule, in the genome-wide (core) model. For such reactions it was possible to compute the RAS.

2.6. Datasets

We applied the MaREA pipeline to two TCGA cohorts.

- The breast cancer dataset (TCGA-BRCA) published in [15], which also includes healthy/control samples. We downloaded the dataset via the cBioPortal [28]. This dataset includes the expression profile (RNA Seq V2 RSEM) of biopsies taken from 817 patients. For 105 of them, the expression profile of the normal tissue is also included.
- The lung adenocarcinoma dataset (TCGA-LUAD, provisional) published in [16], downloaded via the cBioPortal [28]. The dataset includes the expression profile (RNA Seq V2 RSEM) of 586 biopsies from 584 patients. In our analyses we used the data of patients with only one sample.

Because the above datasets identify genes with Entrez IDs, we automatically converted them into *HGNC* IDs. We found a correspondence for 1654 (391) genes over the 1673 (404) included in the genome-wide (core) metabolic model. We opted to neglect missing genes in the computation of the RASs, as done in [6]. However, we were still able to compute a reliable RAS for most reactions associated with a GPR, as missing genes were involved in reactions with an AND operator only in three cases: namely, in the GPRs that involve mitochondrial genes (Complex I, III, IV and V), which are not detected by RNAseq. The MaREA tool provides the user with an alternative option to handle these cases: rules with AND operator, involving missing genes, can be disregarded tout-court.

2.7. Other approaches

As specified in the Introduction (Section 1), metabolic networks are typically used in FBA simulations. In addition to the aforementioned problem regarding the scarce availability of both metabolic and transcriptomic data on the same patient, another challenge when dealing with FBA is the definition of an appropriate objective function.

Although maximization of metabolic growth may approximate well the objective of cancer cells [29], this assumption should not be extended to normal cells, nor generalized to other pathologies. Moreover, integration of transcriptomic data into FBA constraints often prevents the identification of a non-null solution [8,6,7] and requires the *ad hoc* release of some constraints. The impact of this problem is particularly evident when dealing with RNA-seq data, as they are more prone to include null values than transcriptomic data obtained otherwise, e.g., with microarrays, and thus to translate into null constraints. A major example is given by the gene that encodes for the transporter of water from mitochondria to cytosol (AQP8), which may often take value 0. Yet, this reaction is necessary to reach a steady state in which the respiratory chain is used.

Null constraints leading to infeasible solutions may also derive from errors in GPR association rules. Due to their large scope, curation errors are frequent in genome-wide metabolic networks. Genome-wide models are also known to be prone to the presence of thermodynamically infeasible cycles [30], hence they may predict unrealistic flux distributions.

Because of all these reasons, a modeling expert is thus required to guide the biologist among the multitude of current methodologies for transcriptomics integration in constraint-based models, as they are known to provide heterogenous predictions [8,6].

MaREA was conceived to overcome such limitations, providing a simpler framework for trascriptomic data integration in metabolic networks, with relevant translational impact.

The spirit of MaREA is somehow in line with that of Gene Set Enrichment Analysis method (GSEA) [21], which seeks to characterize the sets of up- or down-regulated genes in different phenotypes [21]. The underlying rationale is that even a mild, but concerted, variation in the expression of a set of genes involved in a certain cellular (metabolic)

function might be as relevant as a much larger variation in the individual activity of a single gene.

However, MaREA markedly differs from GSEA. The typical GSEA analysis outcome provides generic indications on the deregulated functions of a cell, or on specific functional behaviors when focusing on particular gene sets, derived, for instance, from the Reactome pathway database [31].

The GSEA enriched sets are list of genes involved in comprehensive, thus potentially quite large, metabolic pathways (as, e.g., "DNA replication"), failing to provide details on which specific metabolic routes are favored in a given condition. In particular, metabolic functions can be alternatively achieved by metabolizing different nutrients and/or by following different catabolic and anabolic routes, in a complex and largely undeciphered interplay. For this reason, simply knowing weather a certain function is up- or down-regulated might not be sufficient to shed light on how such function might be achieved in distinct cancer phenotypes.

More recent approaches aim at building sets of genes to be enriched according to the information included in genome-wide metabolic networks. Specifically, *Metabolic Reporter Analyses* try to provide knowledge about variations in metabolite concentrations, starting from sets of genes that are classified according to the associated metabolites [32]. However, such methods do not provide information about which reactions are up- or down-regulated, and thus hinder the identification of putative targets for cancer treatment.

MaREA can provide a finer resolution to the analysis of metabolism than GSEA and reporter metabolites analyses, by enriching individual metabolic reactions in distinct experimental conditions. Moreover, MaREA improves over current methods based on genome-wide models, by providing a list of curated Gene-Protein-Reaction associations for human central carbon metabolism, and an easy-to-interpret map of corresponding central carbon pathways for visualization of results.

Similarly to the recently introduced PARADIGM approach (PAthway Recognition Algorithm using Data Integration on Genomic Models [33]), MaREA extracts a key metabolic feature (the RAS) for each sample. However PARADIGM relies on integrating data from multiple sources and requires curated pathway interactions among genes, hence both the input data and final objective are significantly different.

As specified above, MaREA can be used to stratify samples in an unsupervised manner. Distinct approaches make use, for instance, of the information on enriched signaling pathways [34] or of that on mutational profiles [35–38] to classify cancer samples and subtypes.

3. Results

3.1. Breast cancer vs. normal (TCGA-BRCA)

In order to evaluate the overall usefulness of MaREA results, we first applied it to a well-known and characterized case-study, the comparison between cancer and normal metabolism. We performed two steps for this analysis: *Reaction Enrichment* and *Reaction Ranking*.

3.1.1. Reaction enrichment

The reactions that have been identified by MaREA as significantly up- or down-regulated in cancer - with at least a 20% increase/decrease - and the magnitude of the deregulation are mapped on the central carbon metabolic network (HMRcore) in Fig. 2, as well as on the genome-wide metabolic network in Supplementary Fig. S1. It can be observed (Fig. 2) that the pathway of glycolysis is over-expressed in cancer. Extensive utilization of glucose is indeed a well established trait of breast and of cancer cells in general [39]. Cancer cells need glucose to feed the metabolic requirement of enhanced proliferation, with particular emphasis on (1) *de novo* synthesis of nucleotides for genome replication; (2) synthesis of amino acids for protein synthesis; (3) synthesis of fatty acids to support the expansion of cellular membranes; (4) ATP generation for energetic requirements.

Accordingly, MaREA returned the following metabolic "modules" as largely up-regulated in cancer: (1) synthesis of nucleotides from Phosphoribosyl-pyrophosphate (PRPP in Fig. 2); (2) metabolism of the non-essential aminoacids serine (Ser), glycine (Gly), alanine (Ala), asparagine (Asn), aspartate (Asp), arginine (Arg) and proline (Pro); (3) synthesis of cholestrol (Chol) from citrate (Cit).

As far as ATP production is concerned, the interpretation of the situation portrayed by MaREA is, as expected, less straightforward. Cancer cells are believed to rely more on fermentation of glucose to lactate rather than on oxidation of glucose in the mitochondria (inner box of the map in Fig. 2), despite the presence of oxygen: a phenomenon well-known as the Warburg Effect [14,13]. However, in contrast to Warburg's original hypothesis that damaged mitochondria are at the root of this phenomenon, the ability of mitochondria to carry out oxidative phosphorylation is not defective in most tumors [13]. In line with these studies, on the one hand lactate secretion seems to be up-regulated in cancer (reaction crossing the external box in Fig. 2). On the other hand, the respiratory chain (represented by the 4 reactions at the bottom of the mitochondrial box, which are catalyzed by protein Complexes I, II, III and IV, plus the reaction that oxidize succinate to fumarate, while reducing ubiquinone to ubiquinol, catalyzed by Complex II) is not significantly down-regulated in this breast cancer dataset, whereas, complex V is slightly up-regulated.

Remarkably, the results shown in Fig. 2 suggest that the working-mode of the TCA cycle may be abnormal in breast cancer, as highlighted in [29]. In particular, up regulation of NADPH-dependent isocitrate dehydrogenase, which catalyzes the reductive carboxylation of α-ketoglutarate (AKG) to isocitrate, may be linked with the mutations often reported for these enzyme in breast and other cancer types [40]. It has been suggested that this enzyme may support reductive glutamine metabolism in cancer and a branched TCA cycle flux mode [29].

The agreement of the results in Fig. 2 with the obvious traits of cancer metabolism supports the reliability of our approach, which might shed light on less established traits. Deregulations of breast cancer metabolism identified by the approach, which may be worth of note are, among others: (1) deregulation of beta-oxidation of palmitate; (2) upregulation of folate metabolism; (3) deregulation of Phosphoenolpyruvate carboxykinase, which converts oxaloacetate (OAA) into phosphoenolpyruvate (PEP).

3.1.2. Reaction ranking

After filtering out the reactions whose activity does not differ between cancer and normal samples, MaREA allows to rank the remaining reactions according to the extent of their up- or down-regulation. Supplementary Table S3 reports the following bits of information for each reaction included in the genome-wide model: the log fold change, the reaction formula, the pathway in which the reaction is involved and a description of its role. As genome-wide models include several reactions that are associated with the very same GPRs, typically involving transporters/enzymes with low substrate specificity, in Fig. 2 we report the top 10 deregulated reactions with different GPRs.

These reactions include 3 up-regulated reactions and 7 down-regulated ones. Most of them are associated with a single gene, including, consistently with the results obtained for the HMRcore model: fatty acids oxidation and PEP carboxykinase, which are significantly down-regulated; lactate (or substrates pertaining to the same family) transport, which is up-regulated in cancer. Single-gene top deregulated reactions not included in the HMRcore model relate to deregulated vitamine A, glycine and alkaloid metabolism and to up-regulated transport of serotonine. Notably, in accordance with this results, it has been reported [41] that serotonine promotes tumor growth and survival in breast cancer, and that vitamin A [42] plays a role in cancer treatment and prevention.

Two top deregulated reactions are associated with a pair of genes linked by an OR and AND respectively: (1) down-regulated antiporter of

A

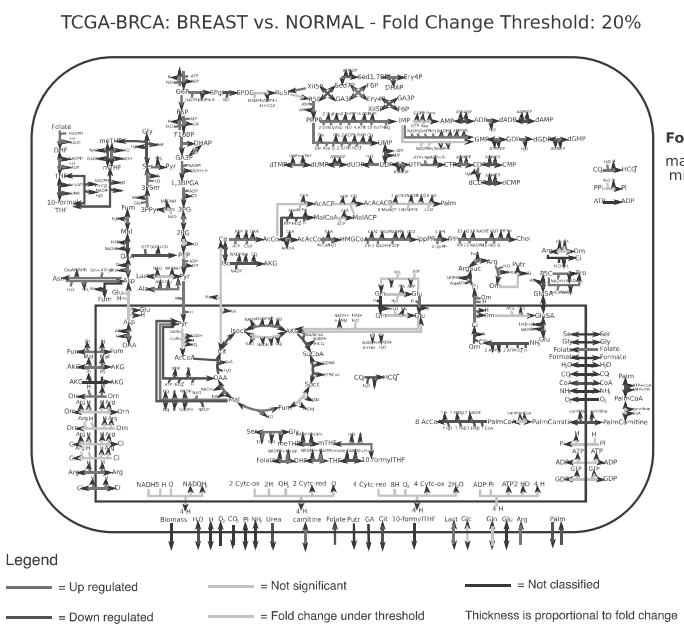


Fig. 2. Reaction enrichment and ranking: breast cancer vs. normal samples. (A) HMRcore map enriched by MaREA: Reactions up-regulated in breast cancer sample set are marked in red, reactions up-regulated in normal sample set are marked in blue. A list of the abbreviations used in the map is provided in the abbreviations section. Thickness of the edges is proportional to the fold-change. Non-Classified reactions, i.e., reactions without information about the corresponding gene-enzyme rule, are marked in black. Dashed gray arrows refer to non-significant deregulations according to the Kolmogorov-Smirnov test. Solid gray arrows refer to reactions with a log2 fold-change below 0.263. (B) A reaction ranking is provided, by listing the 10 reactions with largest log2 fold-change of the RAS (absolute value) in the two conditions. The reaction formula, the corresponding pathway, the gene rule, the up-/down-regulation flag and log2 fold-change are shown in the table. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

B

Reaction	Pathway	Gene Rule	Up-/ Down-Regulated	log2(FC)
3mb2coa[m] + gly[m] <=> CE2026[m] + coa[m] + h[m]	Glycine, serine, alanine and threonine metabolism	HGNC:13734	Down	-5,31
2 k[c] + 2 na1[e] + srtn[e] <=> 2 k[e] + 2 na1[c] + srtn[c]	Transport, extracellular (serotonine)	HGNC:11050	Up	4,95
glyc3p[c] + nad[c] <=> dhap[c] + h[c] + nadh[c]	Glycerophospholipid metabolism	HGNC:4455	Down	-4,85
gtp[c] + oaa[c] -> co2[c] + gdp[c] + pep[c]	Glycolysis/gluconeogenesis (PEP carboxykinase)	HGNC:8724	Down	-4,76
ala_D[e] + gln_L[c] <=> ala_D[c] + gln_L[e]	Transport, extracellular (alanine, serine, glycine, threonine)	HGNC:11026 and HGNC:11058	Down	-4,75
for[c] + nad[c] -> co2[c] + nadh[c]	Fatty acid oxidation	HGNC:3978	Down	-4,70
nad[c] + retinol_9_cis[c] <=> h[c] + nadh[c] + retinal_cis_9[c]	Vitamin A metabolism	HGNC:9940	Down	-4,40
lac_L[e] + na1[e] <=> lac_L[c] + na1[c]	Transport, extracellular (lactate, acetate,	HGNC:19119	Up	4,39
coumarin[c] + h[c] + nadph[c] + o2[c] -> h2o[c] + hcoumarin[c] + nadp[c]	Cytochrome metabolism	HGNC:2608 or HGNC:2610	Up	4,37
egme[r] + h2o[r] -> ecgon[r] + h[r] + meoh[r]	Alkaloid synthesis	HGNC:17	Down	-4,24

the aminoacids alanine, serine, glycine and threonine with glutamine; (2) up-regulated Cytochrome P450 2A6, which is involved in the metabolism of many xenobiotics.

The down-regulation observed for the former reaction is in line with recent studies that have linked the resistance of specific cancer cell lines to amino acid analogs anticancer drugs to a decreased expression of the corresponding transporter [43,44].

The up-regulation identified for the latter reaction (P450 2A enzyme) is worth of note, as P450 enzymes may be involved in carcinogens activation in breast cancer. Environmental carcinogens have been identified in the etiology of breast cancer. For example, CYP2A6 protein detected in the breast can activate nitrosamines and food mutagens to their ultimate carcinogens and thus could play a role in the initiation of breast cancer [45]. Moreover, this enzyme can metabolize clinically important drugs, such as the tamoxifen [46], which represents the most widely used hormonal therapy for breast cancer, and the coumarin [47,48], whose metabolism was proven to produce some metabolites having estrogenic and cytotoxic activities.

3.1.3. Comparison with GSEA results

The GSEA and MaREA approaches are not directly comparable, as they present several differences in goals, input data, parameters, variables and outputs. For instance, MaREA computes an individual activity score for each sample, whereas GSEA only considers expression fold-changes between pairs of experimental conditions. In order to provide an overview of how the information produced by the two complementary approaches may differ, without any claim about which approach should be preferred, we disregarded the addition multiple test correction (FDR) used by GSEA, and we considered the gene-sets that pass the nominal p-value test, with the same threshold used in MaREA standard settings (i.e., $p = 0.05$). We did not set any threshold on the minimum size of gene-sets.

To run GSEA we used two kind of gene sets: (1) curated gene sets based on Reactome, as directly provided by the GSEA tool; (2) gene-sets reconstructed by us, which correspond to the genes associated with each reaction in the genome-wide model and which are provided in *gmt* format (GSEA compliant) in Supplementary File S2.

The first kind of gene-sets represent a typical application of GSEA to gene-sets involved in broad metabolic functions. The second kind is directly comparable to the sets used to compute the RAS by our approach. It should be mentioned that the second type includes many single gene-sets (i.e., size 1) because many reactions are catalyzed by

enzymes associated with a single gene. For this reason, we set the minimum set size to 1 in the GSEA options.

The application of GSEA to Reactome gene-sets returned 144 gene sets significantly enriched in cancer, and 60 gene-sets significantly enriched in normal, at nominal p-value <0.05. When ranking the obtained gene-sets according to the returned Enrichment Score (ES), we observed that, as expected, the first 10 gene sets enriched in cancer refer to generic metabolic functions (in particular: cell cycle and mitosis, asparagine glycosylation, DNA replication and chromosome maintenance, HIV infection and kinesins). The results highlight how MaREA should be used as a complement to GSEA analysis, in order to provide a more fine-grained analysis of metabolic deregulations.

Conversely, the application of GSEA to the more fine-grained datasets, based on Recon 2.2 reactions, returned a number of reactions significantly deregulated much lower than that returned by MaREA. MaREA returned 3339 reactions as significantly up- or down-regulated by at least 20% (p-value <0.05), whereas GSEA returned 105 gene sets significantly enriched in cancer, and 110 gene-sets significantly enriched in normal, at nominal p-value <0.05. This discrepancy is mainly due to the presence of gene-sets including a single-gene, which are reasonably penalized by the GSEA approach. For instance, the single-gene sets associated with serotonin and vitamin A metabolism, which were ranked in the top 10 deregulated reactions by MaREA and might play a role in cancer according to literature, do not pass the nominal p-value test in GSEA. It is worth noticing that also the top-ranked reactions in MaREA that involve genes in OR (Cytochrome metabolism) or in AND (extracellular transport of alanine, serine, glycine and threonine) do not pass the significance test in GSEA.

Taken together, these results indicate that MaREA provides a more complete and refined portray of metabolic deregulations. Moreover, as opposed to GSEA, MaREA computes an independent score for each sample (the RAS), which can be used to cluster samples in an unsupervised fashion. We illustrate such application of MaREA in the next section.

3.2. Sample stratification via MaREA

MaREA can also be used to stratify samples into distinct metabolic subgroups or clusters, also when the presence and number of such clusters is unknown. To this end, in order to provide an experimental validation of the stratification results, we also employed other known measures, such as, e.g., the *survival probability* of patients.

To provide an example application and compare different clustering features, we performed unsupervised clustering on two distinct cancer datasets: (1) the TCGA-BRCA dataset on breast cancer and (2) the TCGA-LUAD dataset on lung adenocarcinoma (see Section 2.6, Datasets, for details).

In particular, we performed a k-means unsupervised clustering over two distinct metabolic networks as input – (1) HMRCore and (2) Recon 2.2 – with different inputs $k \in \{1, 2, ..., 9\}$. We repeated the clustering using three different distance measures, respectively based on: *(a)* normalized RAS,[2] *(b)* RNA-seq data of all metabolic genes, *(c)* the predicted fluxes.

To predict fluxes via FBA computation, we used the E-Flux method [49], which uses the RAS to set constraints on the flux boundaries, and we optimized for growth. The RAS was first normalized across patients as proposed in [50]. We allowed each extracellular nutrient considered in the baseline version of Recon 2.2 to be uptaken/secreted (upper and lower bound set to -1 and 1 respectively, as in [49]).

For each case, we performed $n = 100$ bootstrap iterations, with random centroid assignments, selecting as optimal the clustering run displaying the maximum inter-cluster distance. We then tested the resulting sample clusters against the survival probability (as retrieved from clinical data in the original datasets [15,16]), via a log-rank test on the Kaplan-Meier curves with respect to *overall survival* (OS), *disease-free interval* (DFI), and *disease-specific survival* (DSS).

3.2.1. Metabolic subgroups of breast cancer (TCGA-BRCA)

By applying MaREA to the TCGA-BRCA dataset, we found statistically significant differences ($p < 0.05$) in the Kaplan-Meier curves in the following cases:

(i–a) HMRcore using RAS and $k = 2$, for OS, DFI, and DSS.
(ii–a) Recon 2.2 using RAS and $k = 2$ for DFI.
(i–b) HMRcore using RNA-seq of metabolic genes

and $k = 2$ for DFI, and DSS; and $k = 9$ for OS, DFI, DSS.

All the other cases displayed not significant differences in survival expectancy ($p > 0.05$), including those based on clusters identified on fluxes.

In Fig. 3 we show the most significative result, obtained with (i–a) HMRcore with RAS with $k = 2$ (Subgroup 1: 630/817 samples, Subgroup 2: 187/817 samples). One can see that the curves of the two clusters never overlap, leading to a significant log-rank test ($p = 0.046$).

This result indicates that the up-/down-regulation patterns, as encoded by the RAS values, might indeed be used to split samples in metabolic groups with significantly different prognosis. It is also worth noticing that, by looking at the composition of the two subgroups with respect to the well-established PAM 50 classification [51] (available for 481 on 817 samples), the subgroup with the worst prognosis (Subgroup 2) is largely constituted (i.e., $\sim$ 70%) by samples belonging to the Basal-like group, which are present in a very small percentage in Subgroup 1 (105 Basal-like samples in 107, i.e., the $\sim$ 98%, belong to Subgroup 2), and show almost no samples from Luminal A subgroup. Subgroup 1, instead, is dominated by Luminal A (200 Luminal A samples in 201, i.e., the $\sim$ 99.5%, belong to Subgroup 1) followed by Luminal B and Her 2 subtypes in different proportions.

This result first suggests that there exists a detectable metabolic *signature* of Basal-like cancer samples, and this is, to the best of our knowledge, a novel result, worth of further investigations. Besides, it is worth noticing that the differences observed at the metabolic level indeed translate into distinct survival probabilities, which standard classification may fail to capture.

More in detail, by looking at the reactions significantly up-/down-regulated with respect to the case Subgroup 1 (best prognosis) vs. Subgroup 2 (worst prognosis) portrayed in Fig. 3 for the core model – and in Supplementary Fig. S2 and Table S4 for the genome-wide model – one can see that many reactions that are enriched in cancer against normal are also enriched in worst against best prognosis, including glycolysis, nucleotide synthesis and serine metabolism. Remarkably some metabolic pathways that are not significantly deregulated in cancer are significantly up-regulated in the worse prognosis subgroup, with particular regard to palmitate biosynthesis.

3.2.2. Metabolic subgroups of lung adenocarcinoma (TCGA-LUAD)

In the case of lung adenocarcinoma, the results of sample stratification via MaREA are even more striking. In Fig. 4 one can see the Kaplan-Meier overall survival curves and the corresponding p-value of the log-rank test, with respect to the stratification obtained in the 6 aforementioned scenarios:

(i–a) HMRcore with RAS.
(i–b) HMRcore with RNA-seq of metabolic genes.
(i–c) HMRcore with fluxes.
(ii–a) Recon 2.2 with RAS.
(ii–b) Recon 2.2 with RNA-seq of metabolic genes.
(ii–c) Recon 2.2 with fluxes.

[2] To avoid possible biases due the differences in RAS range and distribution across reactions, we normalizes the RAS value of each sample by dividing by the maximum RAS for that reaction.

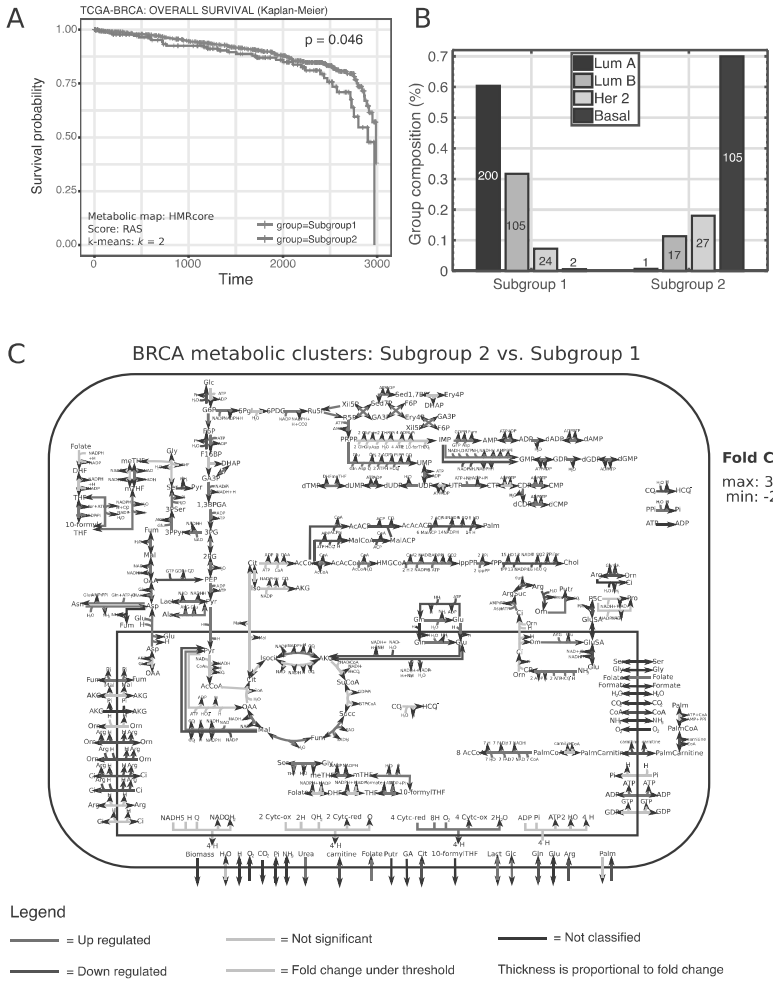


Fig. 3. Breast cancer metabolic clusters. (A) The Kaplan-Meier curves (time unit = days) and the p-value of the log-rank test with respect to the two metabolic clusters of the TCGA-BRCA samples identified by MaREA. The result was obtained using a *k*-means unsupervised clustering ($k = 2$) with HMRcore and RAS. Subgroups 1 and 2 are shown). (B) Composition of the two TCGA-BRCA metabolic clusters identified by MaREA with respect to the PAM50 breast cancer classification. The number of samples in each group is displayed on the bars. (C) Enriched map of HMRcore with respect to the TCGA-BRCA metabolic clusters 1 and 2. A list of the abbreviations used in the map is provided in the Abbreviations Section (see below). Red arrows refer to reactions up-regulated in Subgroup 2 (worst), whereas blue arrows refer to reactions up-regulated in Subgroup 1 (best). Black arrows refer to "Non Classified" reactions, i.e., reactions without information about the corresponding gene-enzyme rule. Dashed gray arrows refer to non significant deregulations according to the Kolmogorov-Smirnov test. Solid gray arrows refer to reactions with a log2 fold change below 0.263.

with respect to $k \in \{1, 2, \ldots, 9\}$; scenarios with $p <= 0.01$ are marked in green, with $0.01 < p <= 0.05$ in yellow, with $p > 0.05$ are not shown.

In all cases, we were able to retrieve clusters with significantly different prognosis for at least some values of *k*, the best results being obtained in cases (*i*–*a*) HMRcore with RAS, (*i*–*b*) HMRcore with RNA-seq of metabolic genes, and (*ii*–*b*) Recon 2.2 with RNA-seq of metabolic genes, which show highly significant p-values ($p \ll 0.01$) for most values of *k* (similar results are obtained for DFI and DSS curves - not shown here).

In this regard, we expect the stratifications with larger *k* to be effective in deciphering the inherent heterogeneity of lung adenocarcinoma, whereas fewer and larger clusters might partially hide this effect. Notice that also in this case the clusters identified on fluxes display the worse predictive power on survival expectancy, at least for higher values of *k*.

Even though the stratification performances of RAS and RNA-seq data of metabolic genes are both remarkable and somehow comparable with respect to survival outcome, we here remark that only with the former approach it is possible to enrich and rank the reactions that distinguish such groups and which, accordingly, might be targeted by metabolic drugs.

For instance, in Fig. 5 we show the metabolic map enriched via MaREA by comparing Subgroup 4 (best prognosis) and Subgroup 3 (worst), in the case of clustering obtained with HMRcore and RAS with $k = 4$ (such stratification provides a good trade-off between a sufficiently large *k* and sufficiently large clusters).

Finally, it is worth noting that some key up-/down-regulation patterns observed when comparing worst versus best prognosis BRCA clusters are conserved in the LUAD case, such as up-regulation of glycolysis, nucleotide synthesis and amino acid metabolism. This result would suggest that key regularities of cancer metabolic deregulation, directly linked with metabolic growth (i.e., DNA and protein synthesis),

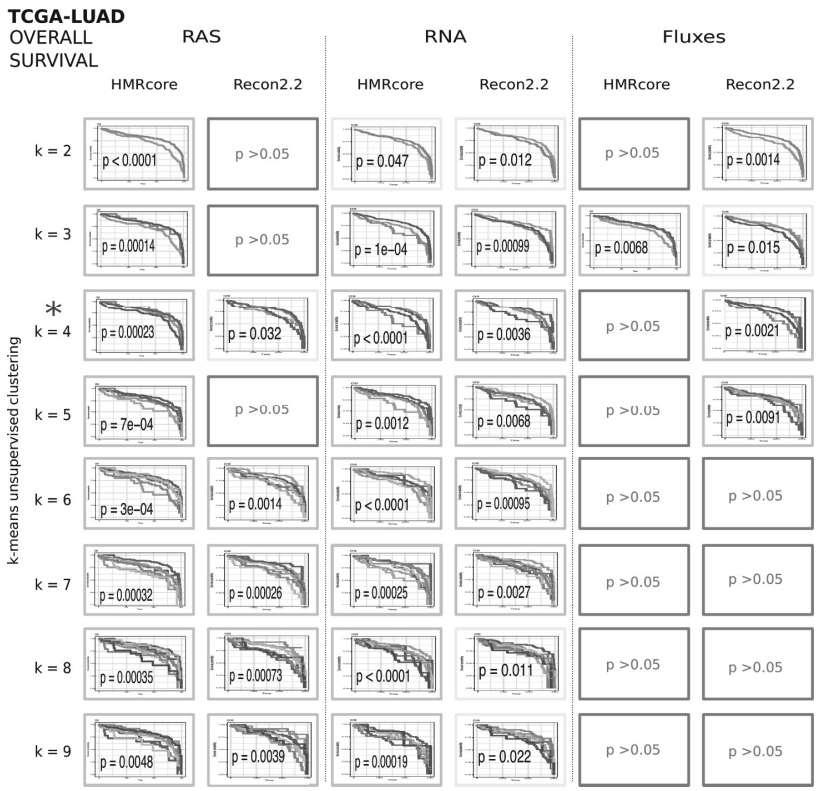


Fig. 4. Survival analysis of lung adenocarcinoma metabolic clusters. The Kaplan-Meier overall survival (OS) curves displaying a p-value of the log-rank test: $p \leqslant 0.01$ (green) or $0.01 < p \leqslant 0.05$ (yellow) are shown, with respect to the *k*-means unsupervised clustering performed on the TCGA-LUAD dataset on breast cancer [16]. 6 cases are considered (from left to right): (*i*—*a*) HMRcore with RAS, (*ii*—*a*) Recon 2.2 with RAS, (*i*—*b*) HMRcore with RNA-seq of metabolic genes, (*ii*—*b*) Recon 2.2 with RNA-seq of metabolic genes, (*i*—*c*) HMRcore with fluxes, (*ii*—*c*) Recon 2.2 with fluxes, for $k \in \{1, 2, \ldots, 9\}$ (time unit = days). The blue star marks the case used to enrich the map in Fig. 5, i.e., $k = 4$, HMRcore with RAS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

might be shared across distinct cancer types, and deserves further investigations.

4. Conclusions

We have here introduced MaREA, a computational pipeline that integrates transcriptomic data into metabolic networks, to compare the metabolism of samples in distinct subgroups or different experimental conditions.

In particular, the reaction enrichment performed via MaREA allows to identify the metabolic patterns underlying the phenotypic and functional properties observed in different sample subgroups, as in the case of (but not limited to) patients with distinct cancer subtypes. This is an important advantage of MaREA, especially when the estimation of metabolic fluxes is not possible or is scarcely reliable. The interpretation of the results is then favored thanks to the effective visualization of up-/down-regulated reactions directly on the metabolic networks.

Furthermore, MaREA can effectively stratify samples in clusters with similar metabolic activity, in unsupervised fashion. Its prognostic power can be evaluated via standard survival analysis.

The case studies on TCGA cancer datasets proved that MaREA can reproduce known properties and traits of metabolic networks in different scenarios. For instance, MaREA allowed to identify the key metabolic paths that distinguish normal from tumor samples, but it also provided cues to formulate and test new experimental hypotheses. Moreover, the metabolic clusters of samples identified by MaREA displayed significantly different survival expectancy, as retrieved from clinical data, for both breast and lung cancer, and this proved that metabolic reprogramming plays a crucial role in cancer aggressiveness.

In the Lung Adenocarcinoma case study (TCGA-LUAD dataset), up to 9 sample subgroups with significantly different prognosis have been identified, by exclusively taking into account transcriptional regulation of metabolic reactions. In the breast cancer case study (TCGA-BRCA dataset), the metabolic differences in terms of patient prognosis were less relevant, suggesting that metabolic heterogeneity might play a milder role in breast cancer as compared to lung cancer. However, MaREA was able to identify two BRCA metabolic clusters characterized by significantly different prognosis and that show a striking overlap with Luminal A and Basal-like standard signatures.

When comparing distinct clustering features, RASs and transcripts proved to be more effective than fluxes in identifying sample subgroups with different prognosis, probably due to the lack of proper constraints on extracellular fluxes in FBA computation.

Besides, clustering based on the core model displayed a better prognostic power than that based on the genome-wide model, at least when employing the RAS, despite a dramatic reduction of the dataset dimensionality. This result highlights the importance of the curation process of the GPR associations that was performed for the core model,

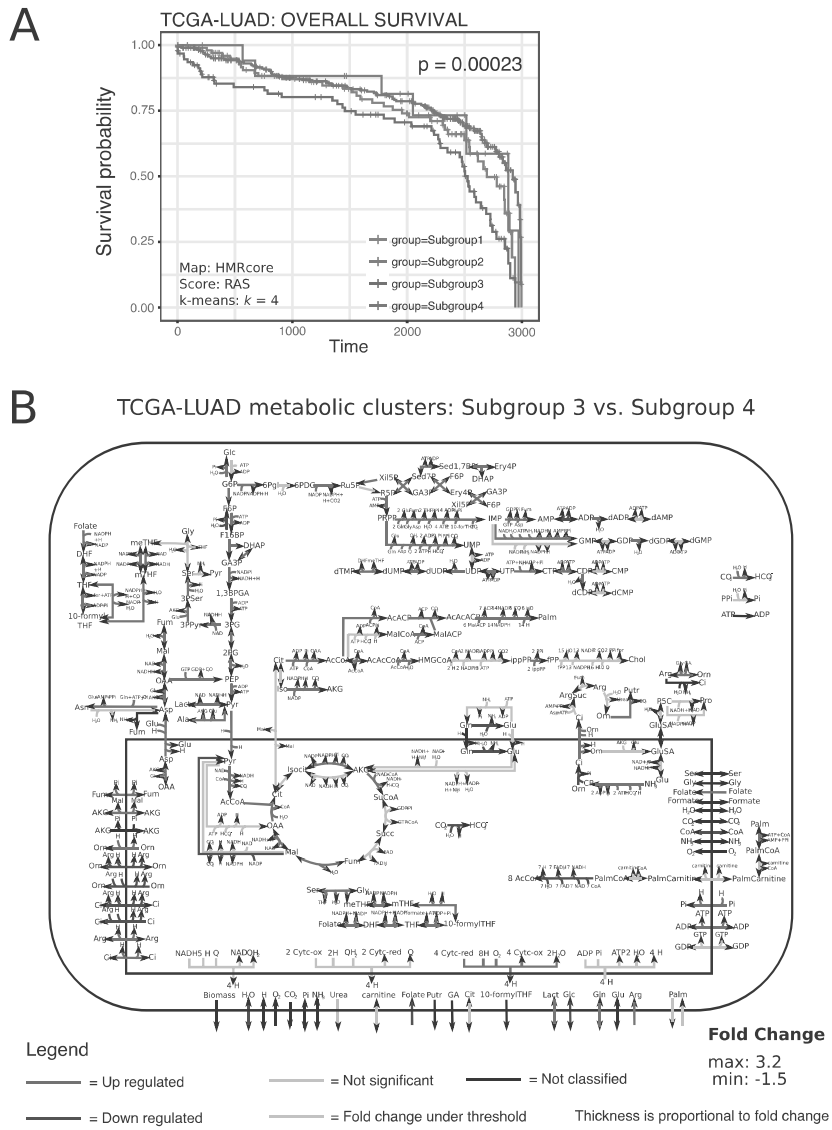


Fig. 5. Lung adenocarcinoma metabolic clusters. (A) The Kaplan-Meier curves (time unit = days) and the p-value of the log-rank test with respect to the two metabolic clusters of TCGA-LUAD samples identified by MaREA with *k*-means unsupervised clustering ($k = 4$) using HMRcore and RAS are shown. (B) Enriched HMRcore map with respect to metabolic clusters 4 (best prognosis) and 3 (worst prognosis), identified by *k*-means unsupervised clustering ($k = 4$) with HMRcore map and RAS. The list of the abbreviations used in the map is provided in Abbreviations Section. Red arrows refer to reactions up-regulated in Subgroup 3 (worst), whereas blue arrows refer to reactions upregulated in Subgroup 4 (best) Black arrows refer to "Not Classified" reactions, i.e., reactions without information about the corresponding gene-enzyme rule. Dashed gray arrows refer to non significant deregulations according Kolmogorov-Smirnov test. Solid gray arrows refer to reactions with a log2 fold change below 0.263.

and which could be further refined.

We also recall that a major benefit of using the RAS, rather than only the transcripts, lays in its effectiveness in summarizing metabolic deregulation when comparing two cohorts or experimental settings. In fact, MaREA allows to identify, rank and visualize critical metabolic reactions, which might more easily be targeted, e.g., with metabolic drugs, rather than targeting the expression of individual genes.

However, as MaREA does not provide information on metabolic fluxes, we believe that MaREA results should be complemented with metabolic measurements and flux simulations, when possible, to provide an all-encompassing picture of metabolism and its deregulation.

Abbreviations

1,3BPGA	1,3-Bisphospho-D-glycerate
10-formyl-THF	10-Formyltetrahydrofolate
2PG	2-phospho-D-glycerate
3PG	3-phospho-D-glycerate
3PPyr	3-phosphonooxypyruvate
3PSer	3-phosphoserine
6PDG	6-phospho-D-gluconate
6Pgl	glucono-1,5-lactone-6-phosphate
AcAcACP	acetoacetyl-ACP

AcAcCoA	acetoacetyl-CoA
AcACP	acetyl-ACP
AcCoA	acetyl-CoA
ACP	acyl carrier protein
ADP	adenosine 5′-diphosphate
AKG	alpha-ketoglutarate
Ala	alanine
AMP	adenosine 5′-monophosphate
Arg	arginine
ArgSuc	argininosuccinate
Asn	asparagine
Asp	aspartate
ATP	adenosine 5′-triphosphate
BRCA	Breast Cancer
CDP	cytidine 5′-diphosphate
Chol	cholesterol
Ci	citrulline
Cit	citrate
CMP	cytidine 5′-monophosphate
CO2	carbon dioxide
CoA	coenzyme A
CP	carbamoyl-phosphate
CTP	cytidine 5′-triphosphate
Cytc-ox	ferricytochrome c
Cytc-red	ferrocytochrome c
dADP	2′-deoxyadenosine 5′-diphosphate
dAMP	2′-deoxyadenosine 5′-monophosphate
dCDP	2′-deoxycytosine 5′-diphosphate
dCMP	2′-deoxycytosine 5′-monophosphate
dGDP	2′-deoxyguanosine 5′-diphosphate
dGMP	2′-deoxyguanosine 5′-monophosphate
DHAP	dihydroxyacetone phosphate
DHF	7,8-Dihydrofolate
dTMP	2′-Deoxythymidine-5′-monophosphate
dUDP	2′-Deoxyuridine-5′-diphosphate
dUMP	2′-Deoxyuridine-5′-monophosphate
Ery4P	erythrose-4-phosphate
F16BP	fructose 1,6-bisphosphate
F6P	fructose 6-phosphate
FAD	flavin adenine dinucleotide
FADH2	flavin adenine dinucleotide reduced
FBA	Flux Balance Analysis
FDR	False Discovery Rate
for	formate
fPP	farnesyl diphosphate
Fum	fumarate
G6P	glucose 6-phosphate
GA	guanidinoacetate
GA3P	glyceraldehyde 3-phosphate
GDP	guanosine 5′-diphosphate
Glc	glucose
Gln	glutamine
Glu	glutamate
GluSA	glutamate 5-semialdehyde
Gly	glycine
GMP	guanosine 5′-monophosphate
GSEA	Gene Set Enrichment Analysis
GPR	Gene-Protein-Reaction
GTP	guanosine 5′-triphosphate
H	proton
H2O	water
HCO3-	hydrogencarbonate
HGNC	HUGO Gene Nomenclature Committee
HMGCoA	hydroxymethylglutaryl-CoA
IMP	2′-inosine-5′-monophosphate
ippPP	isopentenyl diphosphate
Iso	isocitrate
Isocit	isocitrate
KS	Kolmogorov-Smirnov
Lact	lactate
Mal	malate
MalACP	malonyl-ACP
MalCoA	malonyl-CoA
MaREA	Metabolic Reaction Enrichment Analysis
meTHF	5,10-Methenyltetrahydrofolate
mTHF	5,10-Methylenetetrahydrofolate
NAD	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide reduced
NADP	nicotinamide adenine dinucleotide phosphate
NADPH	nicotinamide adenine dinucleotide phosphate reduced
NH3	ammonia
O2	oxygen
OAA	oxaloacetate
Orn	ornithine
P5C	1-pyrroline-5-carboxylate
Palm	palmitate
PalmCarnitine	palmitoylcarnitine
PalmCoA	palmitoyl-CoA
PARADIGM	PAthway Recognition Algorithm using Data Integration on Genomic Models
PEP	phosphoenolpyruvate
Pi	phosphate
Ppi	diphosphate
Pro	proline
PRPP	phosphoribosyl pyrophosphate
Putr	putrescine
Pyr	pyruvate
Q	ubiquinone
QH2	ubiquinol
R5P	ribose 5-phosphate
RAS	Reaction Activity Score
RPKM	Reads per Kilobase per Million mapped reads
Ru5P	ribulose 5-phosphate
SBML	Systems Biology Markup Language
Sed1,7BP	sedoheptulose 1,7-bisphosphate
Sed7P	sedoheptulose 7-phosphate
Ser	serine
Succ	succinate
SuCoA	succinyl-CoA
TCGA	The Cancer Genome Atlas
THF	5,6,7,8-Tetrahydrofolate
UDP	uridine 5′-diphosphate
UMP	uridine 5′-monophosphate
UTP	uridine 5′-triphosphate
VMH	Virtual Metabolic Human
Xil5P	xylulose-5-phosphate

Availability of data and material

MaREA is provided as Matlab tool and is freely available at this link: https://github.com/BIMIB-DISCo/MaREA, along with the source code and the input models. The datasets to reproduce the case studies presented in this paper can be downloaded at this link. The TCGA-BRCA

and TCGA-LUAD datasets used in this work are available for download via the cBioPortal for Cancer Genomics http://www.cbioportal.org/.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported with FOE funds from the Italian Ministry of Education, Universities and Research (MIUR) to SYSBIO - within the Italian Roadmap for ESFRI Research Infrastructures; by grants from AIRC (investigator grant IG 16819 and 9970–2010 Special Program Molecular Clinical Oncology 5x1000) to EM; and by funds from the FLAG-ERA grant ITFoC to GM. The funding sources had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author's contributions

AG, DM and CD designed and implemented the MaREA pipeline; AG, DM, CI, MDF, and CD analyzed the data and applied the MaREA pipeline; DM, MG and MDF curated the model; DM released the MATLAB tool; all the authors analyzed the results and interpreted the models; AG, DM, MA and CD wrote the original draft of the manuscript, which all the authors reviewed and revised in the final form. CD supervised the project.

Acknowledgments

We warmly thank Dario Pescini and Riccardo Colombo for useful discussions and for support to data visualization.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbi.2018.09.010.

References

[1] G.S. Hotamisligil, Inflammation and metabolic disorders, Nature 444 (7121) (2006) 860.
[2] P. Cazzaniga, C. Damiani, D. Besozzi, R. Colombo, M.S. Nobile, D. Gaglio, D. Pescini, S. Molinari, G. Mauri, L. Alberghina, et al., Computational strategies for a system-level understanding of metabolism, Metabolites 4 (4) (2014) 1034–1087.
[3] A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen, J. Nielsen, Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease, Nat. Commun. 5 (2014).
[4] N. Swainston, K. Smallbone, H. Hefzi, P.D. Dobson, J. Brewer, M. Hanscho, D.C. Zielinski, K.S. Ang, N.J. Gardiner, J.M. Gutierrez, et al., Recon 2.2: from reconstruction to model of human metabolism, Metabolomics 12 (7) (2016) 1–7.
[5] I. Thiele, N. Swainston, R.M. Fleming, A. Hoppe, S. Sahoo, M.K. Aurich, H. Haraldsdottir, M.L. Mo, O. Rolfsson, M.D. Stobbe, et al., A community-driven global reconstruction of human metabolism, Nat. Biotechnol. 31 (5) (2013) 419.
[6] D. Machado, M. Herrgård, Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism, PLoS Comput. Biol. 10 (4) (2014) e1003580.
[7] K. Yizhak, B. Chaneton, E. Gottlieb, E. Ruppin, Modeling cancer metabolism on a genome scale, Mol. Syst. Biol. 11 (6) (2015) 817.
[8] S. Opdam, A. Richelle, B. Kellman, S. Li, D.C. Zielinski, N.E. Lewis, A systematic evaluation of methods for tailoring genome-scale metabolic models, Cell Syst. 4 (3) (2017) 318–329.
[9] J.D. Orth, I. Thiele, B.Ø. Palsson, What is flux balance analysis? Nat. Biotechnol. 28 (3) (2010) 245–248.
[10] J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J.M. Stuart, C.G.A.R. Network, et al., The cancer genome atlas pan-cancer analysis project, Nat. Genet. 45 (10) (2013) 1113–1120.
[11] V. Leoncikas, H. Wu, L.T. Ward, A.M. Kierzek, N.J. Plant, Generation of 2,000 breast cancer metabolic landscapes reveals a poor prognosis group with active serotonin production, Sci. Rep. 6 (2016).
[12] N. McGranahan, C. Swanton, Biological and therapeutic impact of intratumor heterogeneity in cancer evolution, Cancer Cell 27 (1) (2015) 15–26.
[13] J.R. Cantor, D.M. Sabatini, Cancer cell metabolism: one hallmark, many faces, Cancer Discov. 2 (10) (2012) 881–898.
[14] P.S. Ward, C.B. Thompson, Metabolic reprogramming: a cancer hallmark even warburg did not anticipate, Cancer Cell 21 (3) (2012) 297–308.
[15] G. Ciriello, M.L. Gatza, A.H. Beck, M.D. Wilkerson, S.K. Rhie, A. Pastore, H. Zhang, M. McLellan, C. Yau, C. Kandoth, et al., Comprehensive molecular portraits of invasive lobular breast cancer, Cell 163 (2) (2015) 506–519.
[16] The Cancer Genome Atlas Network, et al., Comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (7511) (2014) 543.
[17] C. Damiani, M. Di Filippo, D. Pescini, D. Maspero, R. Colombo, G. Mauri, popfba: tackling intratumour heterogeneity with flux balance analysis, Bioinformatics. 33 (14) (2017) i311–i318, https://doi.org/10.1093/bioinformatics/btx251.
[18] M. Di Filippo, R. Colombo, C. Damiani, D. Pescini, D. Gaglio, M. Vanoni, L. Alberghina, G. Mauri, Zooming-in on cancer metabolic rewiring with tissue specific constraint-based models, Comput. Biol. Chem. 62 (2016) 60–69.
[19] J. Schellenberger, R. Que, R.M. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, et al., Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0, Nat. Protocols 6 (9) (2011) 1290–1307.
[20] D. Lee, K. Smallbone, W.B. Dunn, E. Murabito, C.L. Winder, D.B. Kell, P. Mendes, N. Swainston, Improving metabolic flux predictions using absolute gene expression data, BMC Syst. Biol. 6 (1) (2012) 73.
[21] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Nat. Acad. Sci. 102 (43) (2005) 15545–15550.
[22] S. Lloyd, Least squares quantization in pcm, IEEE Trans. Inform. Theory 28 (2) (1982) 129–137.
[23] H. Kitano, A. Funahashi, Y. Matsuoka, K. Oda, Using process diagrams for the graphical representation of biological networks, Nat. Biotechnol. 23 (8) (2005).
[24] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, et al., Tissue-based map of the human proteome, Science 347 (6220) (2015) 1260419.
[25] E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, A. Bairoch, Uniprotkb/swiss-prot: the manually annotated section of the uniprot knowledgebase, Methods Mol Biol. 406 (2007) 89–112.
[26] M. Kanehisa, S. Goto, Kegg: kyoto encyclopedia of genes and genomes, Nucl. Acids Res. 28 (1) (2000) 27–30.
[27] K.A. Gray, B. Yates, R.L. Seal, M.W. Wright, E.A. Bruford, Genenames. org: the hgnc resources in 2015, Nucl. Acids Res. 43 (D1) (2014) D1079–D1085.
[28] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401–404.
[29] C. Damiani, R. Colombo, D. Gaglio, F. Mastroianni, D. Pescini, H.V. Westerhoff, G. Mauri, M. Vanoni, L. Alberghina, A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth: The Warburq effect, PLOS Comput. Biol. 13 (9) (2017) e1005758.
[30] D. De Martino, F. Capuani, M. Mori, A. De Martino, E. Marinari, Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks, Metabolites 3 (4) (2013) 946–966.
[31] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. Gopinath, G. Wu, L. Matthews, et al., Reactome: a knowledge-base of biological pathways, Nucl. Acids Res. 33 (Suppl. 1) (2005) D428–D432.
[32] K.R. Patil, J. Nielsen, Uncovering transcriptional regulation of metabolism by using metabolic network topology, Proc. Natl. Acad. Sci. U.S.A. 102 (8) (2005) 2685–2689.
[33] C.J. Vaske, S.C. Benz, J.Z. Sanborn, D. Earl, C. Szeto, J. Zhu, D. Haussler, J.M. Stuart, Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using paradigm, Bioinformatics 26 (12) (2010) i237–i245.
[34] A. Graudenzi, C. Cava, G. Bertoli, B. Fromm, K. Flatmark, G. Mauri, I. Castiglioni, Pathway-based classification of breast cancer subtypes, Front. Biosci. (Landmark edition) 22 (2017) 1697.
[35] M. Hofree, J.P. Shen, H. Carter, A. Gross, T. Ideker, Network-based stratification of tumor mutations, Nat. Methods 10 (11) (2013) 1108–1115.
[36] D. Ramazzotti, G. Caravagna, L. Olde Loohuis, A. Graudenzi, I. Korsunsky, G. Mauri, M. Antoniotti, B. Mishra, Capri: efficient inference of cancer progression models from cross-sectional data, Bioinformatics 31 (18) (2015) 3016–3026.
[37] G. Caravagna, A. Graudenzi, D. Ramazzotti, R. Sanz-Pamplona, L. De Sano, G. Mauri, V. Moreno, M. Antoniotti, B. Mishra, Algorithmic methods to infer the evolutionary trajectories in cancer progression, Proc. Nat. Acad. Sci. 113 (28) (2016) E4025–E4034, https://doi.org/10.1073/pnas.1520213113.
[38] G. Caravagna, Y. Giarratano, D. Ramazzotti, I. Tomlinson, T.A. Graham, G. Sanguinetti, A. Sottoriva, Detecting repeated cancer evolution from multi-region tumor sequencing data, Nat. Methods 15 (9) (2018) 707.
[39] R.A. Gatenby, R.J. Gillies, Why do cancers have high aerobic glycolysis? Nat. Rev. Cancer 4 (11) (2004) 891–899.
[40] Z.J. Reitman, H. Yan, Isocitrate dehydrogenase 1 and 2 mutations in cancer: alterations at a crossroads of cellular metabolism, J. Natl Cancer Inst. 102 (13) (2010) 932–941.
[41] V.P. Pai, A.M. Marshall, L.L. Hernandez, A.R. Buckley, N.D. Horseman, Altered serotonin physiology in human breast cancers favors paradoxical growth and cell survival, Breast Cancer Res. 11 (6) (2009) R81.
[42] E. Doldo, G. Costanza, S. Agostinelli, C. Tarquini, A. Ferlosio, G. Arcuri, D. Passeri, M.G. Scioli, A. Orlandi, Vitamin a, cancer treatment and prevention: the new role of cellular retinol binding proteins, BioMed Res. Int. 2015, Article ID 624627, 14 pages. http://dx.doi.org/10.1155/2015/624627.

[43] V. Ganapathy, M. Thangaraju, P.D. Prasad, Nutrient transporters in cancer: relevance to warburg hypothesis and beyond, Pharmacol. Therapeut. 121 (1) (2009) 29–40.
[44] Y. Huang, W. Sadée, Membrane transporters and channels in chemoresistance and-sensitivity of tumor cells, Cancer Lett. 239 (2) (2006) 168–182.
[45] H. Hellmold, T. Rylander, M. Magnusson, E. Reihnér, M. Warner, J.-A. Gustafsson, Characterization of cytochrome p450 enzymes in human breast tissue from reduction mammaplasties, J. Clin. Endocrinol. Metabol. 83 (3) (1998) 886–895.
[46] I. Bièche, I. Girault, E. Urbain, S. Tozlu, R. Lidereau, Relationship between intratumoral expression of genes coding for xenobiotic-metabolizing enzymes and benefit from adjuvant tamoxifen in estrogen receptor alpha-positive postmenopausal breast carcinoma, Breast Cancer Res. 6 (3) (2004) R252.
[47] C. Rodriguez-Antona, M. Ingelman-Sundberg, Cytochrome p450 pharmacogenetics and cancer, Oncogene 25 (11) (2006) 1679–1691.
[48] M.A. Musa, J.S. Cooperwood, M.O.F. Khan, A review of coumarin derivatives in pharmacotherapy of breast cancer, Curr. Med. Chem. 15 (26) (2008) 2664–2679.
[49] C. Colijn, A. Brandes, J. Zucker, D.S. Lun, B. Weiner, M.R. Farhat, T.-Y. Cheng, D.B. Moody, M. Murray, J.E. Galagan, Interpreting expression data with metabolic flux models: predicting mycobacterium tuberculosis mycolic acid production, PLOS Comput. Biol. 5 (8) (2009) e1000489.
[50] A. Brandes, D.S. Lun, K. Ip, J. Zucker, C. Colijn, B. Weiner, J.E. Galagan, Inferring carbon sources from gene expression profiles using metabolic flux models, PLoS One 7 (5) (2012) e36947.
[51] M.C. Liu, B.N. Pitcher, E.R. Mardis, S.R. Davies, P.N. Friedman, J.E. Snider, T.L. Vickery, J.P. Reed, K. DeSchryver, B. Singh, et al., Pam50 gene signatures and breast cancer prognosis with adjuvant anthracycline-and taxane-based chemotherapy: correlative analysis of c9741 (alliance), NPJ Breast Cancer 2 (2016) 15023.