UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
The genomic landscape of beneficial fitness effects

Permalink
https://escholarship.org/uc/item/0g19v9w9

Author
Hsu, Brian

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

The genomic landscape of beneficial fitness effects

A thesis submitted in partial satisfaction of the requirements for the degree
Master of Science

in

Bioengineering

by

Brian Bowen Hsu

Committee in charge:
      Trey Ideker, Chair
      Anne Carvunis
      Shankar Subramanium
      Kun Zhang

2016

The Thesis of Brian Bowen Hsu is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Chair

University of California, San Diego

2016

Dedication

To my family, thank you for your continual support throughout the years.

Dad, thank you for all the care and love you have given me since I came into this world. From the bottom of my heart, I can say that I wouldn't be the man I am today if it wasn't for you. I swear I will find a way to repay you one day.

Grandma, your life is a true inspiration to me. Your perseverance through all the challenges you faced in your life reminds me to never give up.

David, thank you for always being there for me and supporting me in all my endeavors. I may not show it at times but I truly appreciate all you have done for me. I will give back to you everything you have done for me one day.

Gran, may you rest in peace. Thank you for loving me as one of your own.

Grandpa, rest in peace. I wish you could have seen me today, but don't worry I'll tell you all about it one day.

Epigraph

Nothing in biology makes sense except in the light of evolution.

—*Christian Theodosius Dobzhansky (1973)*

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge Professor Ideker for his support as the chair of my committee. I would also like to acknowledge Anne for all the support and guidance she has provided me.

I would also like to acknowledge the "Yeast team", who spent tireless hours performing all of the wet lab experiment. Without their support, I would not have been able to complete my research.

Chapter 2, in part is currently being prepared for submission for publication of the material. Hsu, Brian; Medetgul-Ernar, Kate; Hines, Cameron; Michaca, Manuel; Regent, Nick; Carvunis, Anne; Ideker, Trey. The thesis author was the primary investigator and author of this paper.

Chapter 3, in part is currently being prepared for submission for publication of the material. Hsu, Brian; Medetgul-Ernar, Kate; Hines, Cameron; Michaca, Manuel; Regent, Nick; Carvunis, Anne; Ideker, Trey. The thesis author was the primary investigator and author of this paper.

Chapter 4, in part is currently being prepared for submission for publication of the material. Hsu, Brian; Medetgul-Ernar, Kate; Hines, Cameron; Michaca, Manuel; Regent, Nick; Carvunis, Anne; Ideker, Trey. The thesis author was the primary investigator and author of this paper.

ABSTRACT OF THE THESIS

The genomic landscape of beneficial fitness effects

by

Brian Bowen Hsu

Master of Science in Bioengineering

University of California, San Diego, 2016

Professor Trey Ideker, Chair

In this thesis, I develop a high-throughput framework to systematically identify beneficial genetic perturbations relative to a wild-type strain using an ultra-high density colony array. I applied the method to measure the fitness of ~8000 deletion and overexpression mutants across 6 metabolic conditions after growth on either an environment supplemented with all amino acids or an amino acid limited environment. I identified a set of genes corresponding to the respiration pathway, specifically the pyruvate dehydrogenase (PDH) complex, whose loss of function increases fitness. This phenomenon is involved in cancer pathogenesis and contributes to the Warburg effect. We find that the beneficial growth effect of loss of PDH function is only present

under amino acid limited environments, which could suggest conditional requirements for beneficial events in cancer.

I also found that the environmental history experienced by the cell affects the process of adaptation to a novel environment and can lead to a beneficial growth effect. Although this is a transient process, it is further evidence for the role of epigenetic inheritance in determining evolutionary outcome.

Chapter 1: Introduction

Organisms are constantly exposed to new environments and must adapt to each successive environment in order to survive. Short-term survival strategies include regulation of gene expression to produce the necessary proteins required for survival in a new environment[1–3]. Long-term and more permanent strategies include accumulation and propagation of beneficial mutations, a process known as adaptive evolution[4,5]. Currently, not much empirical evidence exists about beneficial mutations. However, understanding how beneficial mutations manifest in the genome will not only provide insights into the origin of biological mechanisms but also into key phenotypes of medical relevance such as drug resistance in cancer and bacteria[6,7]. To understand this phenomenon, a promising approach would be to systematically probe the entire genomic landscape and identify potential drivers of adaptation.

Recent technological advances in next-generation sequencing (NGS) have made it feasible to identify beneficial mutations in evolve and resequence (E&R) experiments[8,9]. E&R experiments use laboratory evolution to adapt populations to a specific environment, followed by NGS to analyze the resulting genetic changes. This approach can be used to uncover the molecular determinants of adaptation. However, these experiments are resource intensive—requiring many replicates of a single strain to achieve sufficient sequence coverage to call a mutation, time consuming—lasting from

weeks to months, and expensive—requiring multiple reagents and sequencing runs. Most E&R experiments are conducted in bacteria and yeast, in part, because of their short generation time; performing the same experiment in mammalian cells could take 10x longer due to the difference in generation time. Experimental design and execution is not the only challenge in these experiments; analysis of the resulting mutation profile to differentiate adaptive from passenger mutations can also be challenging.

All of the studies to date performing E&R experiments have been conducted in static environmental conditions[9–11]. While the results of these experiments provide deep insights into adaptive evolution they do not accurately reflect changes that would occur in nature, following a succession of different environmental conditions. Adding this degree of complexity to a laboratory evolution experiment would further decrease the tractability of the experiment.

On the other hand, genome-wide screening technology has existed for over a decade and helped functionally annotate most genes by identifying specific phenotypes, the most common ones being slow growth and cell death[12]. Genome-wide screens typically utilize a library of mutants, where each mutant contains a single gene perturbation. The *S. cerevisiae* Deletion Project lead to the creation of the Yeast Knockout (YKO) collection, the first and only complete, systematically constructed deletion collection for any organism[13,14]. The YKO collection has been used in a wide array of genome-wide phenotypic screens aimed at understanding biological functions, stress

response, and mechanism of drug action. Historically, the majority of these screens have focused on characterizing deleterious genetic perturbations, which demonstrate a relationship between a gene and survival in an environment. There remains a lack of research focused on beneficial genetic perturbations because of their rarity. It was also recently discovered that beneficial genetic perturbations identified in genome-wide screens have a high concordance with beneficial mutations identified in E&R experiments[11]. This opens up the possibilities of taking advantage of the high-throughput, low-cost, and easy-to-use nature of traditional screening methodologies to study adaptive evolution.

Chapter 2. Framework to identify beneficial genetic perturbations

2.1 Introduction

The use of genome-wide screens has been paramount in systematically identifying the genotype-to-phenotype relationship in model organisms including bacteria, yeast, and humans. This has been made possible through the utilization of genome-wide libraries, an experimentally tractable tool constructed from systematically perturbing most genes in the genome[13,15,16]. The quantitative readout of these experiments is an estimate of fitness, which is the overall contribution to survival and reproduction. Fitness is estimated for each mutant containing a perturbed gene during growth in a specific environmental condition to identify the functional relationship between gene and environment.

Genome-wide screens in *S. Cerevisiae* measuring fitness fall into three main categories: (1) colony size measurement on solid media[17], (2) growth rate estimation in liquid media[18,19], and (3) barcode counts after competitive growth in liquid media[20,21]. The effect of the genetic perturbation is then determined based on the fitness of the altered mutant relative to the entire population. The major limitation of growth rate estimation in liquid media is the difficulty to increase throughput. The most common plates used are 384-density plates. Therefore, sixteen 384-density plates would be required to screen the entire yeast genome at one replicate per gene on a single condition. This method would not be feasible because one replicate will not

4

provide sufficient statistical power to detect significant changes and screening multiple conditions would be extremely resource intensive. The highest density plate available is 1536-density plates, however these plates are expensive due to their intricate design. On the other hand, the development of NGS has solved this problem and significantly increased the throughput of barcode-sequencing based methods. However, the main pitfall with barcode-sequencing is the competition between mutants, which could confound the interpretation of the results.

Because most screens have pinned cells on solid media, bioinformatics packages to analyze this data type have been extensively developed to generate robust and reproducible measurements of fitness based on colony size[17,22]. These efforts are reflected by the highest average cross-study correlation for colony-size based fitness measurements compared to fitness measurements in competitive and liquid growth, suggesting that colony-size captures fitness at a higher resolution and with less noise[17].

In the majority of these studies the phenotypic readout is either slow growth or cell death, leading to the identification of genes with a deleterious effect when perturbed. These methods have resulted in the identification of thousands of genotype-to-phenotype relationships. However, traditional methods to identify beneficial genetic perturbations are less well developed. Beneficial genetic perturbations are defined as mutants whose fitness is higher than the wild-type (WT) strain. Few studies exist that have attempted to measure fitness of these perturbations using traditional screening methods

(Table 1). However, some of these studies neglect to use a control strain making it difficult to estimate the true null distribution of WT behavior resulting in a higher false positive rate[11,23]. Other studies potentially grew their cultures for too many generations[24] increasing the likelihood of secondary mutations based on an estimated mutation rate of ~10^{-10}/bp/generation[25,26], a phenomenon known to occur with the yeast deletion collection[27]. All of the previous studies have also measured fitness following competition in liquid media, where cells are known to cooperate with one another[28,29]. These cell-cell interactions will likely lead to different resultant phenotypes compared to isolated growth, which may explain the ~25-50% correlation between fitness estimations in solid media and competition in liquid[17]. The competitive and cooperative interactions between strains could potentially confound the identification of true beneficial genetic perturbations.

Here, we have developed an experimental and computational framework to identify beneficial genetic perturbations while circumventing the experimental limitations of previous attempts. We built a high-density colony array using the genome-wide yeast knockout collection and a single wild-type strain interspersed throughout the array in order to quantify the fitness effects of beneficial genetic perturbations in *S. Cerevisiae*.

Table 1: Compilation of studies characterizing beneficial genetic perturbations in S. cerevisiae. Few studies exist that have systematically studied beneficial genetic perturbations. Most of the previous studies used a liquid-based approach, while we used an agar-based approach.

References	Species	Ploidy	Fitness measurement	Growth	Mutants	Control strain	Conditions tested
Sliwa et al. 2005	*S. Cerevisiae*	Homozygous diploid	Serial dilution	Liquid	~4000	HO deletion	N = 1
Delneri et al. 2008	*S. Cerevisiae*	Heterozygous diploid	Growth rate	Liquid	~4000	All mutants	N = 1
Qian et al. 2012	*S. Cerevisiae*	Homozygous diploid	Barcode-sequencing	Liquid	~4000	HO deletion	N = 6
Payen et al. 2016	*S. Cerevisiae*	Haploid, Homozygous diploid	Barcode-sequencing	Liquid	~8000	All mutants	N = 3
This study	*S. Cerevisiae*	Haploid	Colony size	Agar	~8000	HO deletion	N = 6

2.2 Results

2.2.1 Identification of beneficial genetic perturbations.

To identify beneficial genetic perturbations, we took advantage of the haploid yeast knockout collection containing 4,976 nonessential gene deletions[13]. We used YDL227C (HO) as the WT strain because it is the same background strain as all other deletion mutants tested. The deletion of HO prevents mating type switching and is known to have a neutral effect. We re-arrayed all deletion mutants with the WT strain (N = 3072) in order to minimize plate-to-plate variability when making mutant to WT comparisons **(Figure 2.1)**. The high number of WT replicates increases the resolution of the null distribution and the power to detect changes in fitness. Previous high-throughput screens lacked a WT reference strain and used the weighted

average of the population as the reference[13,30,31]. First, because the underlying null distribution of wild-type behavior is unknown, it is challenging to confidently classify a mutant as either beneficial or deleterious. Second, because the proportion and effect size of deleterious mutants are larger than that of beneficial mutants, the average fitness of the population is lower than the fitness of the WT leading to false classifications of beneficial genetic perturbations.

We measured the colony size of each strain, which was used to estimate fitness. We compared each deletion mutant to the WT strain to determine if its fitness significantly deviates from WT behavior and calculated a Q-value to correct for multiple testing hypothesis. We were able to achieve such a high number of replicates by utilizing 6144-density plating technology, which increased the throughput of our screen 4-fold[32].

We measured the fitness of each deletion mutant on synthetic complete media + galactose (SCG) and minimal media + galactose (SDG) to assess the effects of two commonly used laboratory conditions on yeast growth. Each deletion mutant is classified as beneficial (Q < 1%, fitness > 99.5% of WT fitness), deleterious (Q < 1%, fitness < 0.5% of WT fitness), and neutral based on an effect size threshold of the WT distribution and a Q-value cutoff. The Q-value was used to control the false discovery rate at 1% under multiple testing hypothesis. On SCG, 3% of nonessential gene deletions are beneficial, 18% are deleterious, and 79% are neutral **(Figure 2.2)**. On SDG, 9% of nonessential gene deletions are beneficial, 20% are deleterious, and 71% are

neutral. The number of deleterious mutants are similar to those reported previously[13]. The majority of deletions are neutral, followed by deleterious, and the smallest proportion are beneficial.
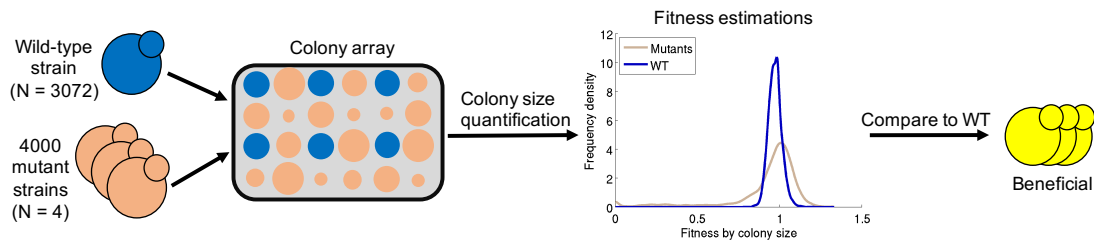
Figure 2.1: Experimental pipeline to identify beneficial genetic perturbations. Wild-type (ΔHO) and 4,976 mutant strains are arrayed on the same plate. Each plate is imaged followed by colony size quantification using "The Colony Analyzer Toolkit" (MATLAB). Fitness is estimated from the normalized colony size. A strain is classified as beneficial if its fitness is significantly larger than wild-type fitness. The distribution of all mutants and the wild-type strain is shown in SDG.
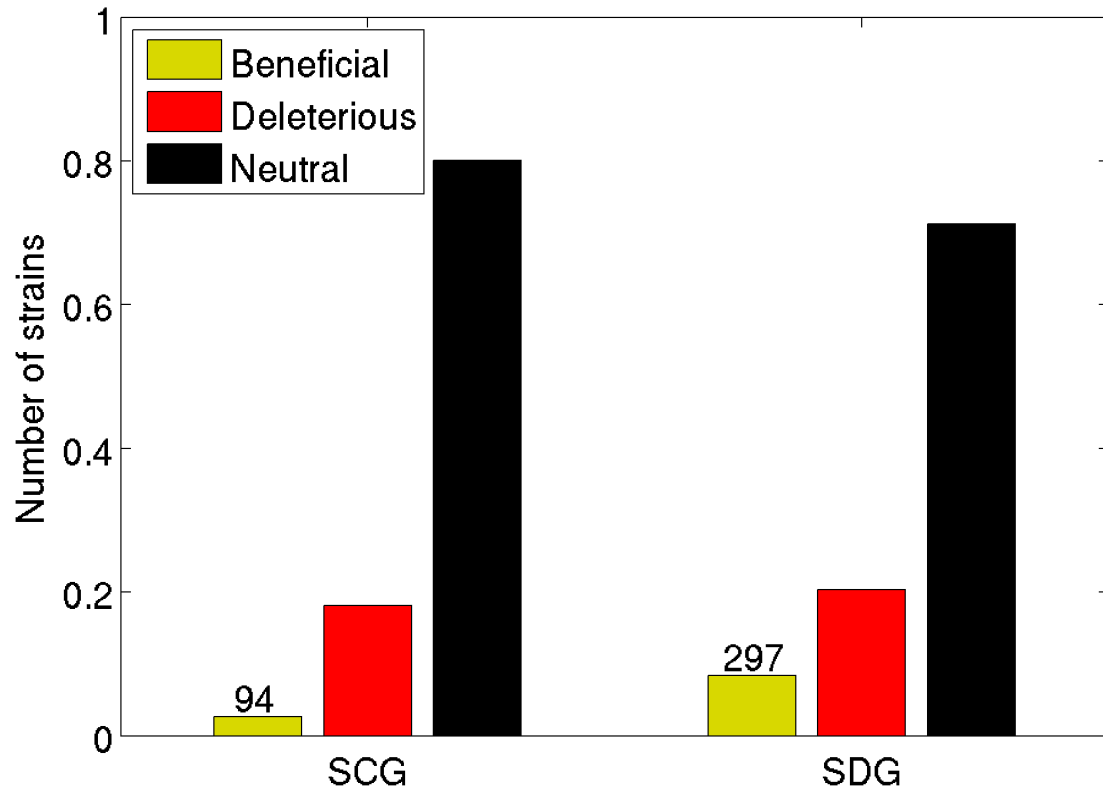
Figure 2.2: Beneficial genetic perturbations are environment-dependent.
94 deletion mutants have a fitness significantly higher than WT in SCG, while
297 deletion mutants have a higher fitness in SDG relative to WT. Deletion of
the majority of genes in the genome result in neutral fitness effects.

2.2.2 Reproducibility of fitness estimations

The reliability of our fitness estimations is reflected by the high
Pearson's correlation coefficient between technical replicates **(r = 0.96; Figure
S2.1A)** and biological replicates **(r = 0.97; Figure S2.1B).** Biological replicates
were completed 1-2 months apart and represent freshly grown cells starting
from the glycerol freezer stock. The similarity between Pearson's correlation
for biological and technical replicates suggest our results are reproducible in
independent experiments.

However, this does not directly reflect the reproducibility of our identified beneficial mutants. To estimate reproducibility, we applied our pipeline to two independent experiments in SDG and calculated the significance of the overlap between the identified beneficial mutants in each experiment. We found 297 and 232 beneficial mutants in each biological replicate with 154 overlapping mutants. We found a statistically significant overlap between the identified beneficial mutants using a hypergeometric test, suggesting our results are reproducible **(P = 9.51 x 10^{-86}; Figure 2.3A)**. We also calculated a reproducibility rate of 41% by taking the ratio between number of overlapping mutants and the number of unique mutants identified in both experiments. To verify the significance in reproducibility, we calculated the expected reproducibility between experiments based on random chance by randomly sampling the number of identified beneficial deletions in each experiment (N = 297 and N = 232) from the population of mutants with a fitness greater than the median WT fitness and calculating the resulting reproducibility rate **(Figure 2.3B)**. The average expected overlap is ~7% while the actual overlap of beneficial mutants is ~41%, verifying the reproducibility of our identified mutants. Although the reproducibility may be slightly low, it is statistically significant and greater than expected by chance. We chose to narrow down the population of mutants to those with a fitness greater than the median fitness of the WT strain to account for the inherent difference in reproducibility between beneficial and deleterious mutants. Deleterious

mutants have a reproducibility rate of ~64% and have a more significant

overlap **(Figure S2.2).**
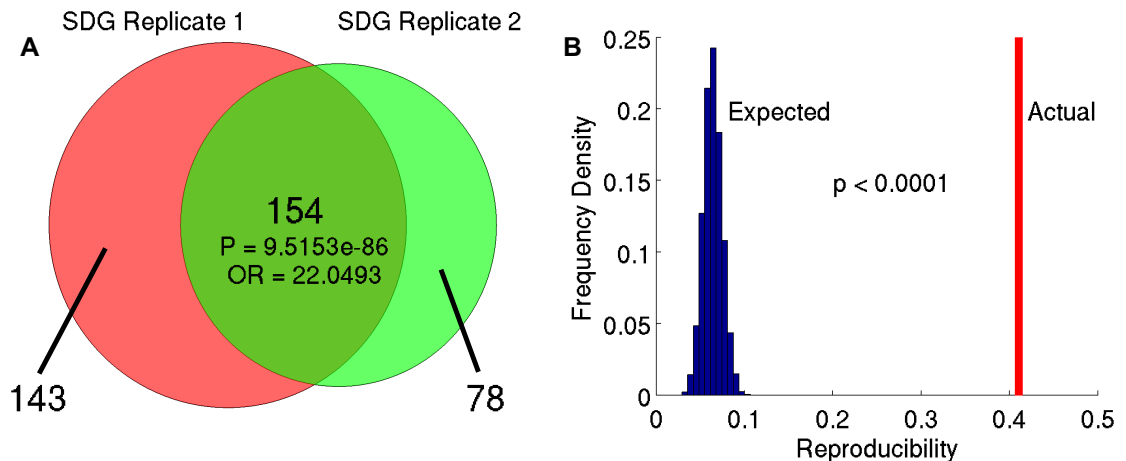
Figure 2.3: Identified beneficial mutants are reproducible in independent experiments on SDG. (A) Beneficial mutants are identified on SDG in two independent experiments. The identified gene sets significantly overlap each other based on a hypergeometric test (P = 9.04 x 10^{-87}). Reproducibility is calculated as the number of overlapping mutants between two sets over the number of unique mutants. **(B)** The expected distribution is calculated by randomly sampling 297 and 232 from mutants with fitness > median WT fitness in each experiment followed by reproducibility calculation.

2.2.3 Comparison to related datasets.

To further understand the reproducibility of our identified beneficial mutants, we compared our results to another study identifying beneficial genetic perturbations using bar-seq based fitness estimations[33]. We tested our pipeline in an independent experiment on YPAD and detected 130 beneficial mutants compared to 227 identified previously resulting in a small but significant overlap of 11 **(P = 0.02; Figure 2.4).** The difference in identified beneficial mutants could be attributed to the difference in background strains tested in each experiment, haploid deletion strains vs homozygous diploid

deletion strains. The correlation between experiments **(r = 0.49; Figure S2.3)** is within the range of correlations (0.14 – 0.70) calculated previously between different methods of fitness estimation using the homozygous diploid strain[17]. However this signal is driven by the deleterious mutants (r = 0.45), while the correlation between beneficial mutants is almost negligible (r = 0.05). Previous experiments found haploid strains to evolve faster and have higher growth rates than diploid strains[34,35], which could explain the larger range of positive fitness effects in our experiment **(Figure S2.3).** Another explanation could be due to the inherent differences between fitness estimations, competitive growth in liquid media followed by bar-seq vs. growth on solid media followed by colony size measurement. The competition between ~4000 mutants could decrease the maximum fitness achievable by a single mutant whereas growth on solid media is less affected by the growth of other mutants.

We applied the same analysis to the deleterious mutants and detected 667 mutants compared to 1270 identified previously resulting in a more significant overlap of 436 **(P = 7.10 x^{-164}, Figure S2.4).** These results are consistent with the difference in calculated reproducibility rates between beneficial and deleterious mutants described above. The distribution of fitness effects for deleterious mutants ranges from 0~0.9 compared to 1.05~1.5 for beneficial mutants. This could suggest larger fitness effects produce more reproducible phenotypes compared to smaller fitness effects, which are prone to more fluctuations in fitness. Fluctuations could be more likely to occur under weaker phenotypes because the selective pressure is not as strong, thus

small changes in fitness will not have an immediate impact on the overall survival of the cell.
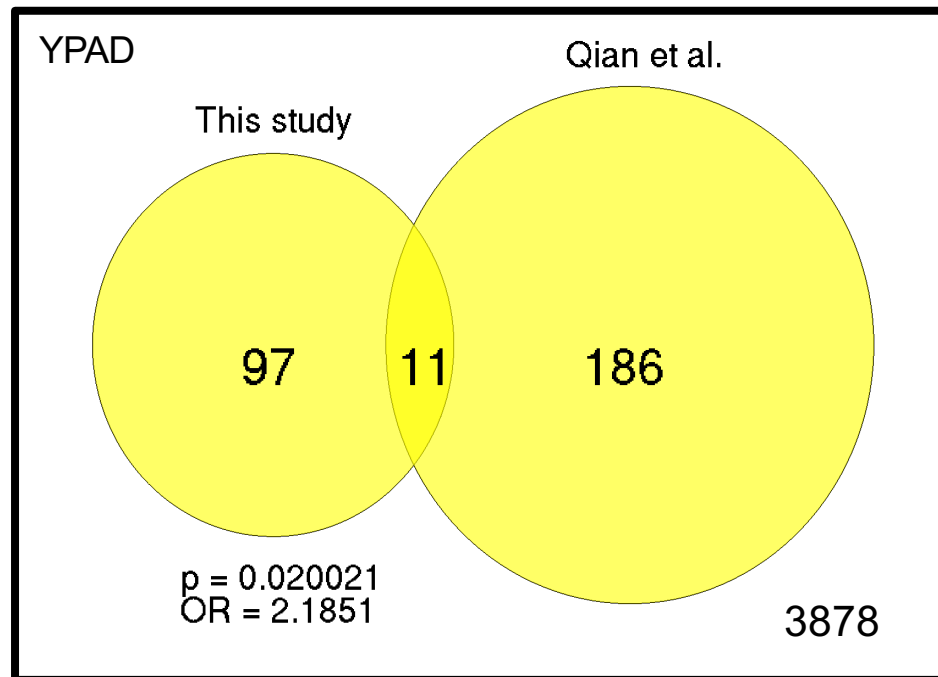
Figure 2.4: Overlap with beneficial mutants identified in previous studies. We identify 130 beneficial deletions vs. 227 in Qian et al. 2012 on YPAD. After intersecting gene spaces, these numbers reduce to 108 and 197, respectively. 11 beneficial mutants are found in both datasets (P = 0.02, OR = 2.18).

2.2.4 Validation of identified beneficial genetic perturbations.

We randomly selected 59 beneficial candidates from SDG and performed an independent experiment using a lower-throughput 1536-density colony array to validate our results. To better estimate the wild-type distribution, we included 5 pseudogene mutants, 14 randomly selected mutants from the barFLEX collection[36] with their plasmids removed using 5-fluoroorotic acid (5-FOA) along with the ΔHO. A pseudogene is a DNA segment that is highly homologous with a functional gene but contains mutations that prevents its expression[37]. Therefore, deletion of these genes is

expected to be neutral. The barFLEX collection contains overexpression mutants using a pGAL-ORF-URA3 plasmid. Removal of these plasmids using 5-FOA will result in wild-type behavior. Because our pipeline uses one biological replicate of WT, however with 3072 technical replicates, we may be underestimating the variance of the true wild-type distribution. We included these 19 biological replicates of strains exhibiting "wild-type" behavior in addition to the ΔHO strain used in the screen to provide a better estimation of the wild-type distribution.

We calculated the validation rate of the candidate beneficial genetic perturbations as the fraction of candidates growing faster than WT over the set of beneficial mutants tested in the validation. We repeated this analysis for the deleterious and neutral mutants. We found that the strains identified as beneficial in out screen are more likely to validate than deleterious or neutral strains. However, when using the same effect size threshold in the validation experiment (99.5% of WT), we calculated a ~18% validation rate **(Figure 2.5A)**. Although this is fairly poor, when the effect size threshold is increased from 3 standard deviations (~99.7% of WT) to 6 standard deviations, the validation rate significantly increases from ~18% to ~78% **(Figure 2.5B)**. One explanation for this discrepancy is the variance of the wild-type distribution in the validation experiment is significantly larger than that of the wild-type distribution in the main experiment. Therefore, the effect size thresholds will be different between the screen and the validation experiment. The co-

occurrence of an increase in validation rate with an increase in effect size

threshold suggests that high effect size mutants display a true biological effect.
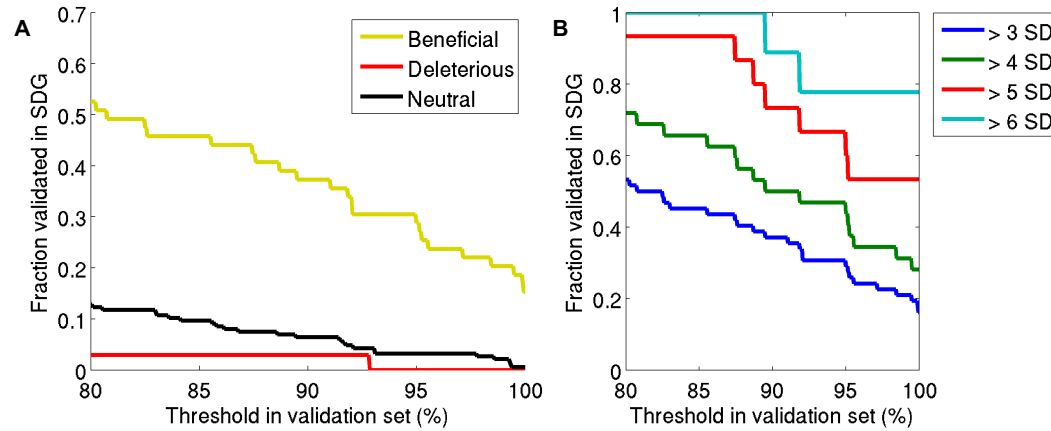
Figure 2.5: Validation rate of identified beneficial mutants in SDG.
Fraction validated is the fraction of mutants identified as beneficial in the
validation experiment over the fraction of beneficial mutants (identified in the
screen) tested in the validation set. **(A)** The fraction validated is highest for
beneficial mutants and there is a 3-4x enrichment between the beneficial and
neutral mutants. **(B)** Fraction validated of beneficial mutants when using
different effect size thresholds in the screen. Effect size thresholds range from
3-6 standard deviations above the mean WT fitness. We find a significantly
higher validation rate when the initial effect size threshold is higher.

2.3 Discussion

In order to create a platform to identify beneficial genetic perturbations,

we developed an experimental and computational framework for measuring

the fitness effects of beneficial mutants using a genome-wide colony array.

Our framework provides a high-throughput and cost-effective alternative to

E&R experiments for studying adaptive evolution. We demonstrate our

platform is able to detect true beneficial mutants through validation in an

independent experiment using a low-throughput colony assay. In the two

environmental conditions tested, we found 3-9% of deletions to provide a

beneficial fitness effect. Our findings are in accordance with the widely held view across multiple evolutionary theories that beneficial mutations are rare and occur less frequently than neutral and deleterious mutations[38,39].

Our framework can also be applicable to different systems such as cell lines. The majority of studies in cell lines are based on negative selection and synthetic lethality screens to identify drug targets[40,41]. However, there remains a fundamental question to be answered: how does cancer evolve? Cancer evolves through driver mutations providing cancerous cells with an increased proliferative ability compared to surrounding non-cancerous cells, resulting in unabated proliferation and eventually tumorigenesis. The key challenge is the identification of driver mutations because they are sparse and vary significantly patient-to-patient. Current solutions involve computational methods leveraging machine-learning algorithms and biological networks such as network-based stratification[42] to analyze data from The Cancer Genome Atlas[43]. We provide an alternative methodology to study cancer evolution, through the identification of adaptive clones providing a significant positive fitness effect under either drug-treated conditions or high-glucose environments. Genome-wide CRISPR libraries have been developed recently[15,16], which can be utilized to screen cancer cell lines and identify the set of deletions conferring the largest positive fitness effects. The identified candidates could then become a list of potential driver mutations.

Chapter 2, in part is currently being prepared for submission for publication of the material. Hsu, Brian; Medetgul-Ernar, Kate; Hines, Cameron;

Michaca, Manuel; Regent, Nick; Carvunis, Anne; Ideker, Trey. The thesis author was the primary investigator and author of this paper.

2.4 Supplementary figures

Table 2: List of media recipes.

Growth Media	Composition (1L)
YPAD	10g yeast extract, 20g bacto peptone, 20g dextrose, 20g bacto agar, 120mg adenine
Synthetic complete + galactose (SCG)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 2g SC dropout mix, 20g galactose
Synthetic complete + casamino acids + galactose (SCCG)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 2g SC dropout mix, 5g casamino acids, 20g galactose
Minimal media + galactose (SDG)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 0.25g histidine, 0.25g uracil, 0.25g methionine, 1.25g leucine, 20g galactose
Synthetic complete + galactose + raffinose (SCGR)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 2g SC dropout mix, 20g galactose, 10g raffinose
Synthetic complete + casamino acids + galactose + raffinose (SCCGR)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 2g SC dropout mix, 5g casamino acids, 20g galactose, 10g raffinose
Minimal media + galactose + raffinose (SDGR)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 0.25g histidine, 0.25g uracil, 0.25g methionine, 1.25g leucine, 20g galactose, 10g raffinose
Minimal media + galactose (SDG_D)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 0.07g histidine, 0.07g uracil, 0.07g methionine, 0.36g leucine, 20g galactose
Minimal media + lysine + galactose (SDLG)	1.7g yeast nitrogen base without amino acids and ammonium sulfate, 1g monosodium glutamic acid, 0.22g histidine, 0.22g uracil, 0.22g methionine, 0.22g lysine, 1.11g leucine, 20g galactose
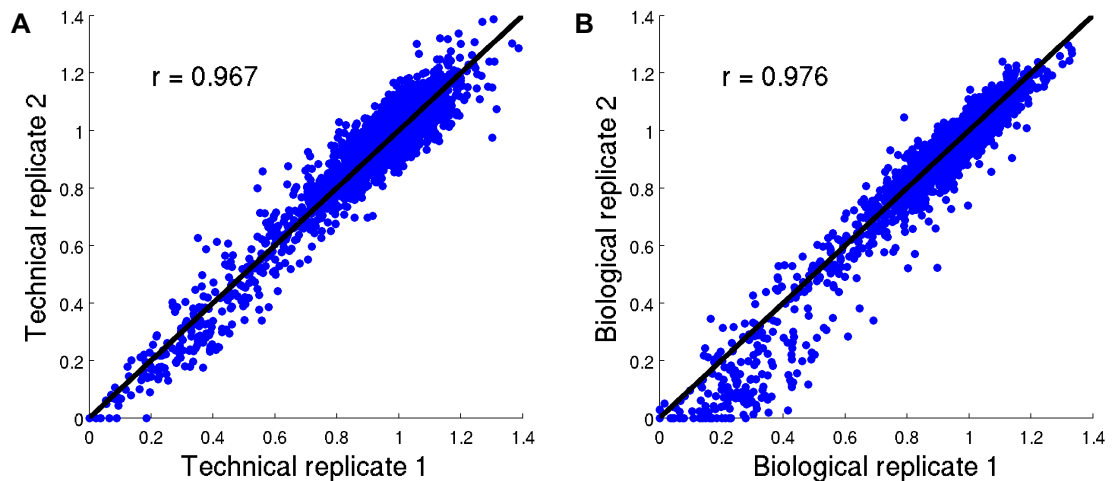
Figure S2.1: Correlation of fitness estimates across replicates. (A) Fitness estimates in SDG are highly correlated between technical replicates (r = 0.967). Technical replicates are mutants from the same starting population grown at the same time. (B) Fitness estimates in SDG are highly correlated

between biological replicates (r = 0.976). Biological replicates were grown fresh from glycerol stocks at different times (months apart).
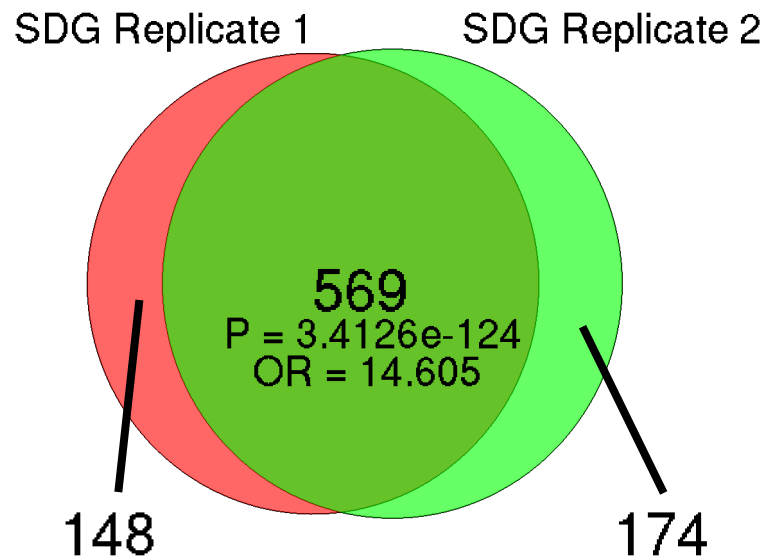
Figure S2.2: Reproducibility of identified deleterious mutants in SDG. Deleterious mutants are identified in independent biological replicates on SDG. Identified deleterious mutants have a significant overlap (P = 3.41 x 10^{-124}) and ~64% reproducibility.
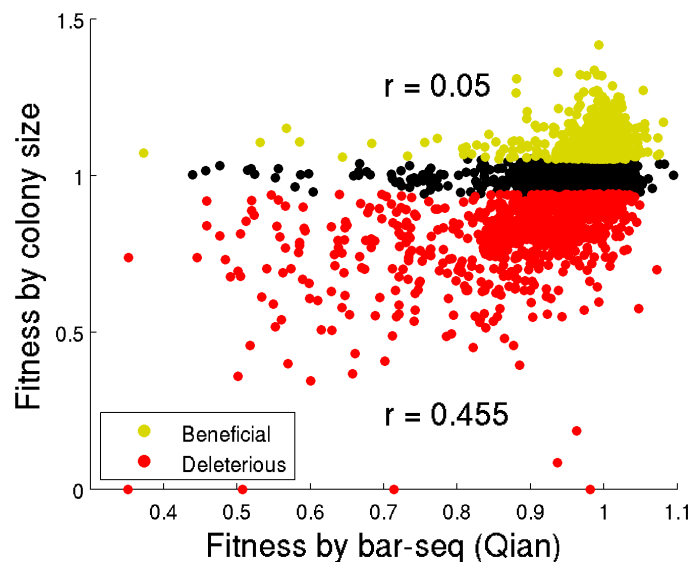
Figure S2.3: Correlation of fitness estimates using colony size vs. bar-seq. Fitness estimates using colony size vs. bar-seq on YPAD are correlated

(r = 0.48). Overall correlation is driven by the deleterious mutants (r = 0.45), while there is no correlation in the beneficial mutants (r = 0.05).
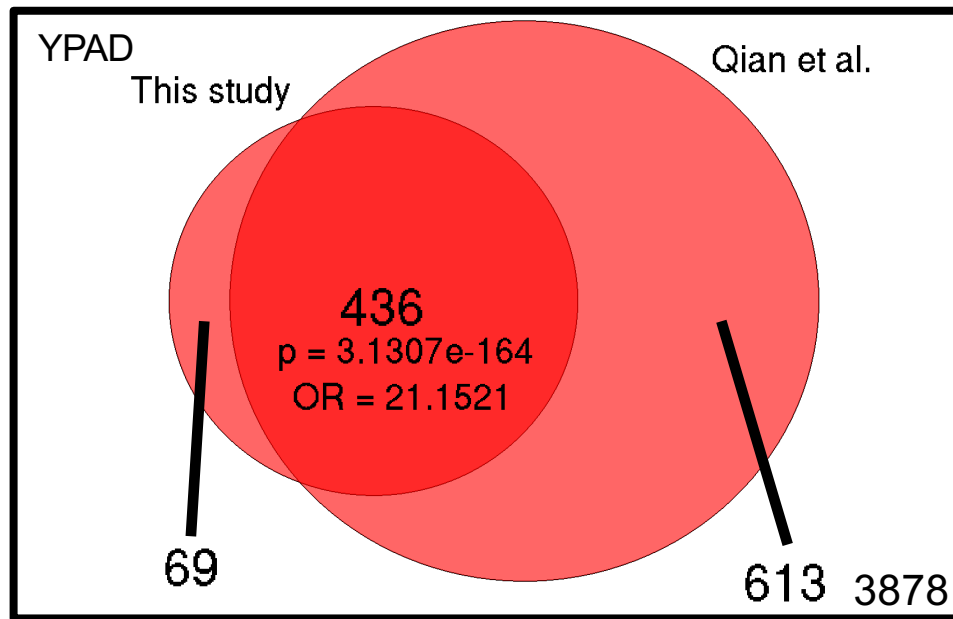
Figure S2.4: Overlap with deleterious mutants identified in previous studies. We identify 667 deleterious deletions vs. 1270 in Qian et al. 2012 on YPAD. After intersecting gene spaces, these numbers reduce to 108 and 197, respectively. 436 beneficial mutants are found in both datasets (P = 3.13 x 10^{-164}, OR = 21.15).

Chapter 3. Genomic landscape of beneficial genetic perturbations

3.1 Introduction

Cell survival is dependent on the coordination of thousands of cellular processes each regulated via a distinct subset of genes. Genes can interact with each other at the DNA, RNA, and protein level. These interactions are responsible for the majority of biological processes to sustain life. The combinations of all of these interactions contributes to the overall phenotype of the cell. Therefore, the relationship between genotype and phenotype is not a simple one-to-one mapping between genes and phenotype. Over the past few decades, our understanding of the genotype-phenotype relationship has evolved to reflect this **(Figure 3.1).** Genes are known to affect two or more distinct and seemingly unrelated traits to produce different phenotypes, an effect known as pleiotropy[33,44,45]. Thus, the same genotype may result in different phenotypes depending on the environment[46]. The effects of these genotype-environment interactions are then propagated along the underlying biological network resulting in a complex phenotype[47].

However, the genotype of a cell can be modified through a vast array of genetic perturbations ranging from point mutations and deletions/insertions to copy number variations and gene duplications. Most perturbations can be categorized as either loss-of-function (LOF), when a gene product has reduced or no function, or gain-of-function (GOF), when an altered gene product possesses a new molecular function or new pattern of gene

expression. Both LOF and GOF mutations can have a variety of effects on the cell ranging from lethality to uncontrolled proliferation. For example, in cancer, GOF mutations in TP53 can inhibit AMPK activation signaling causing upregulation of anabolic processes and glycolytic potential promoting oncogenic functions[48,49]. LOF mutations in BRCA1/BRCA2 lead to impairment in G1/S and G2/M cell cycle checkpoint activation as well as homologous recombination mediated repair of double stranded breaks, which can lead to tumorigenesis[50–52]. Cancerous cells harboring these mutations have an increased proliferative ability allowing them to grow faster and outcompete non-cancerous cells for nutrients. Therefore, it is also important to consider the methods of perturbation (deletion and overexpression) when performing systematic analysis of each gene in order to determine the full spectrum of the genotype-phenotype relationship.

Yeast maintain a species-specific balance in energy production between respiration and fermentation based on either aerobic or anaerobic conditions. Respiration yields 36 ATP and various cofactors such as NADH and $FADH_2$ through the Citric Acid cycle (TCA) and oxidative phosphorylation. This process is more energetically favorable than fermentation, which yields 2 ATP and NADH. Similar to other eukaryotes, some yeast species such as *Kluyveromyces lactis* prefer respiration in aerobic environments with glucose and fermentation in anaerobic environments[53]. However, other species such as *S. cerevisiae* prefer fermentative metabolism even under aerobic conditions and glucose availability, a phenomenon known as the "Crabtree effect"[54]. The

Crabtree effect is mediated by glucose repression, which describes the inducible repression of respiration when glucose is available leading to fermentative metabolism. However, this does not completely eliminate the capacity for respiratory metabolism. As opposed to the "Warburg effect", which describes the irreversible switch to fermentative metabolism due to genetic mutations[55,56]. Thus, the balance in energy production determines the fate of pyruvate, which is the precursor for three major pathways: 1. Gluconeogenesis, 2. Ethanol fermentation, 3. Respiration.

The Warburg effect is a metabolic condition and a defining feature of certain types of cancer cells and has been extensively studied. It provides a growth advantage allowing cancer cells to outgrow normal cells. However, its exact benefits for cell growth and survival are not yet resolved. We hypothesized that the growth advantage of the Warburg effect could be quantified in a systematic screen for beneficial mutants with repressed respiratory metabolism in *S. cerevisiae*. We applied our experimental and computational framework to classify ~8000 engineered *Saccharomyces cerevisiae* deletion and overexpression mutants across 6 metabolic conditions including amino acid limiting environments. We then applied hierarchical clustering on the identified beneficial mutants to understand whether specific biological modules play a role in the Warburg effect and the contribution of each module.
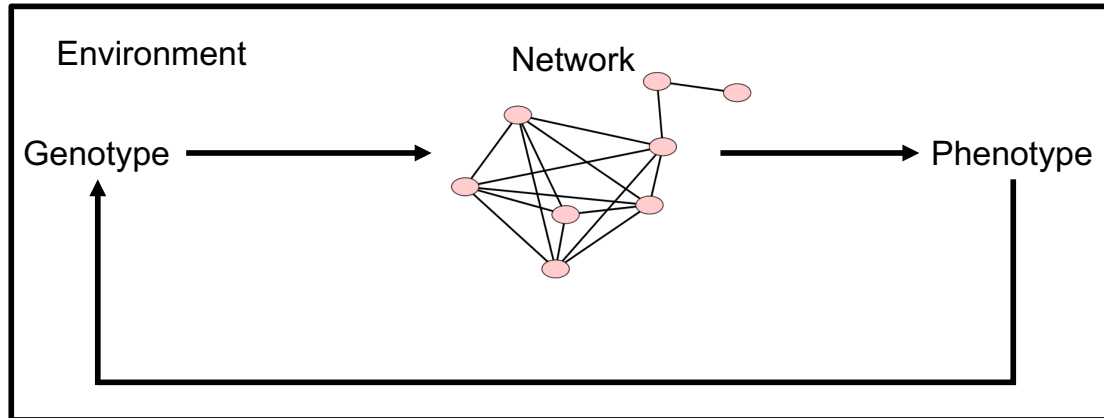
Figure 3.1: Fundamental systems level genotype-phenotype relationship. One theory for the genotype-phenotype relationship is based upon the interaction between genotype, environment, and biological network to produce a phenotype. The phenotype then forms a feedback loop to regulate the genotype for survival.

3.2 Results

3.2.1 Identification of beneficial genetic perturbations across multiple environmental conditions.

We identified beneficial deletion and overexpression mutants from the yeast knockout collection[13] and barFLEX overexpression collection[36] across 6 different environmental conditions consisting of various combinations of carbon and nitrogen sources listed in **Table 2**. We identified >4% of deletions to be beneficial and >15% of deletions to be deleterious in each of the conditions tested **(Figure 3.2)**. For overexpression mutants, <1% are beneficial and ~15% on average are deleterious across all conditions tested. Overexpression mutants have larger experimental variability between conditions tested, which could be because these six experiments were completed at 3 separate times (SCG/SDG/SDGR, SCCG, SCCGR/SDGR).

The first striking result is the >20-fold difference between the fraction of beneficial deletion mutants compared to beneficial overexpression mutants **(Figure 3.2).** Overexpression mutants are constructed using a GAL1/10 promoter to drive expression of each gene on a URA3-based CEN plasmid pBY011[36]. GAL1/10 promoters are induced approximately 1000-fold in the presence of galactose and strongly repressed in glucose[57]. Therefore, there are two explanations that could explain the discrepancy: 1. Overexpression of gene products at high concentrations is mostly either neutral or deleterious, 2. Cells use transcriptional regulation, post-transcriptional modifications, and other regulatory mechanisms to completely reduce the expression of each gene product producing no significant phenotype compared to wild-type. The latter explanation is not likely because we find a 43% overlap (p < 0.0001) between deleterious mutants identified and the set of 769 toxic genes identified in an independent overexpression experiment using the GAL1/10 promoter confirming the overexpression system is functional[58].

The second result is that there is a significantly larger fraction of beneficial deletions identified in SDG and SDGR (N = 425, 453) compared to other conditions tested (N ~ 170) **(Figure 3.2A)**. The increased fraction of beneficial mutants identified in SDG and SDGR is accompanied by a two-fold increase in maximum fitness effect size (**Figure 3.3).** SDG contains the same total amount of supplemented nutrients as SCG, however SDG is only supplemented with histidine, leucine, uracil, and methionine while SCG contains all amino acids and uracil at a lower concentration. Growth on amino

acid limited environments induces the general amino acid control (GAAC) mechanism via activation of GCN4, a transcriptional activator regulating genes expressed during amino acid starvation[59,60]. Deletion of GCN4 decreases fitness to 0.26 in SDG and maintains constant fitness at 1.04 in SCG confirming activation of GCN4 in SDG (**Figure 3.4**). Cells are in different cellular states in SDG compared to SCG; the transcriptome is rewired to upregulate and downregulate the necessary genes for survival in each environment. All of the environments supplemented with amino acids (SCG, SCCG, SCGR, SCCGR) result in similar number of beneficial mutants (N ~ 170), while the amino acid limited environments (SDG, SDGR) result in similar number of beneficial mutants (N ~ 430) **(Figure 3.2A)**. This could suggest that a non-random relationship exists between the number of beneficial mutants identified and the resulting cellular state based on the severity of the environmental stress experienced.
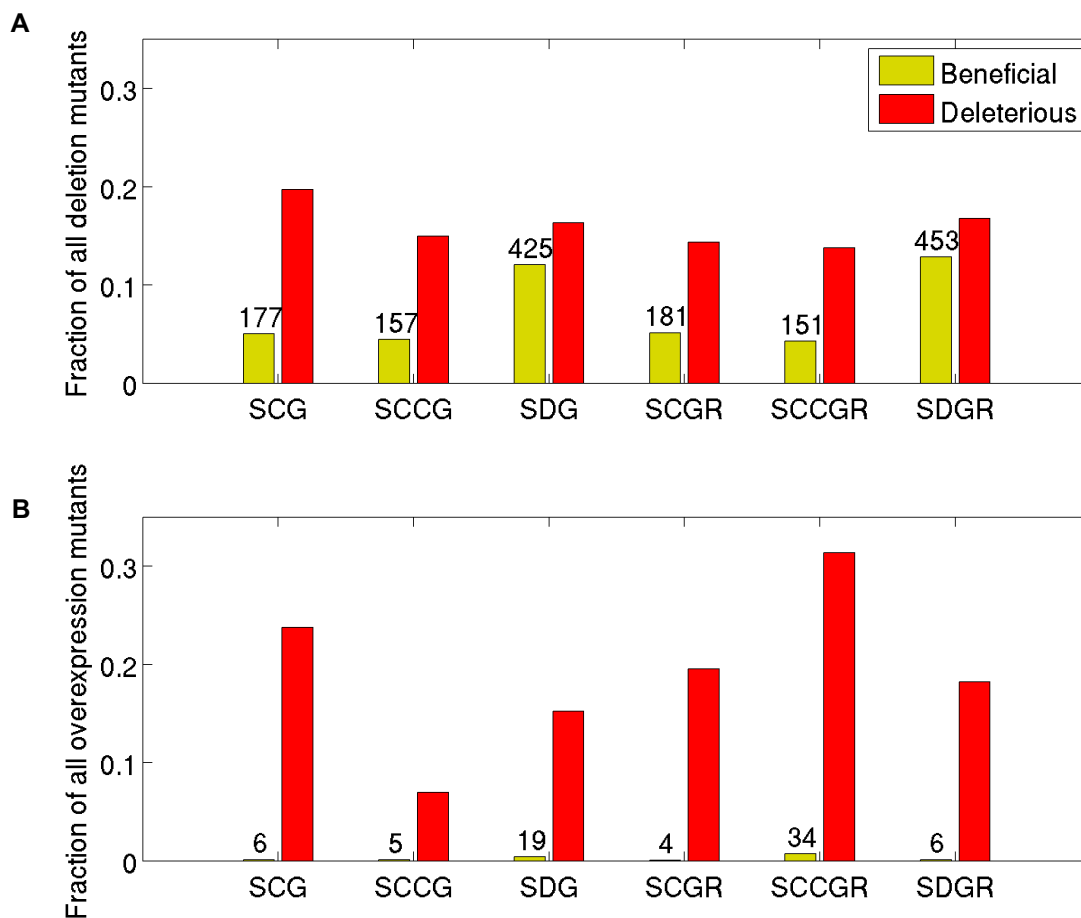
Figure 3.2: Genome-wide identification of beneficial mutants in six environmental conditions. Each bar represents the total number of deletion and overexpression mutants identified as either beneficial or deleterious across six different environmental conditions using the methods described in Chapter 1. The same six conditions were tested for the deletion and overexpression mutants.
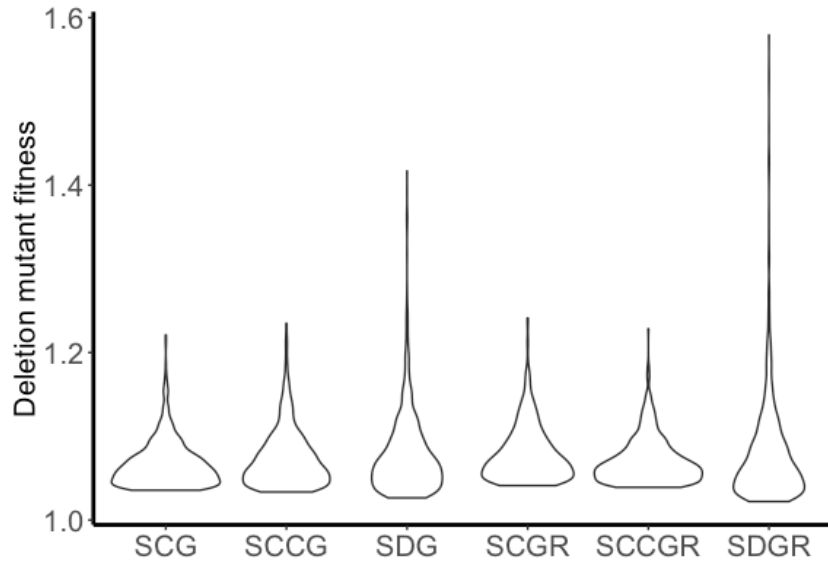
Figure 3.3: Fitness range for identified beneficial deletion mutants. The fitness effects of identified beneficial deletion mutants under an amino acid limiting environment (SDG and SDGR) is between 1 and 1.6, while the range for conditions with all amino acids supplemented is between 1 and 1.25.
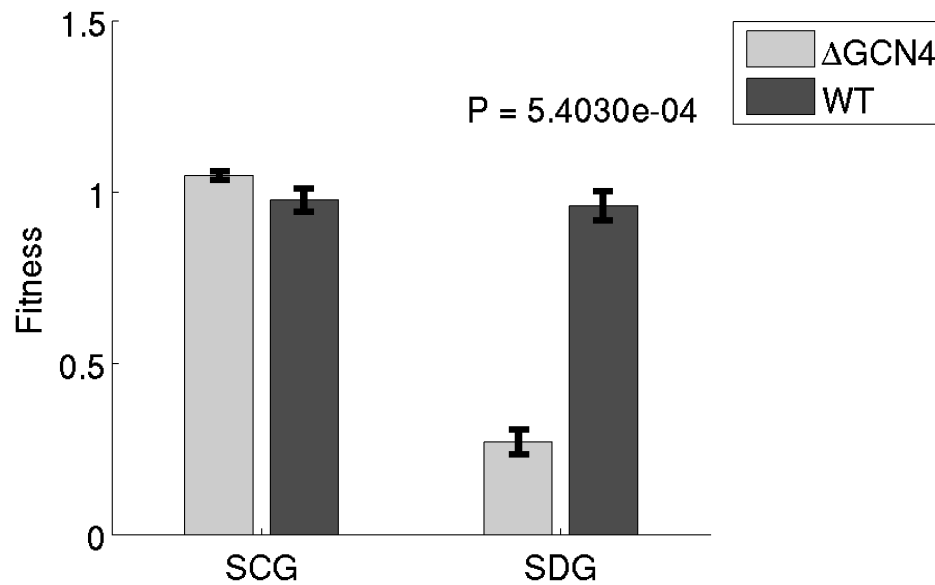
Figure 3.4: Deletion of GCN4 under amino acid limiting environments is deleterious. GCN4 is activated under amino acid limiting environments (SDG and SDGR). Therefore, the deletion of GCN4 causes a significant deleterious fitness effect (**P = 5.4 x 10^{-4}**).

3.2.2 Clustering fitness profiles of beneficial genetic perturbations

We selected the beneficial genetic perturbations with the strongest effect size in each condition using a more stringent filter (Q<0.01, ES>99.9% of WT) for further analysis (N = 289). We correlated the fitness profiles, i.e. the fitness of the deletion or overexpression of a gene across all conditions tested (N = 12), of the selected mutants against each other to generate a correlation matrix. The correlation matrix was hierarchical clustered and the resulting heat map is shown in **(Figure 3.5)**.

Functionally related genes with similar fitness profiles tend to cluster together[31,61]. We performed functional enrichment analysis to identify if clusters of beneficial genes enrich for gene ontology (GO) terms[62]. We found three significant clusters representing genes involved in respiration (P < 0.001), phosphatase activity (P = 0.026), and amino acid metabolic processes (P = 0.035). These results provide further evidence that beneficial mutations are not random and specific biological modules are targeted.
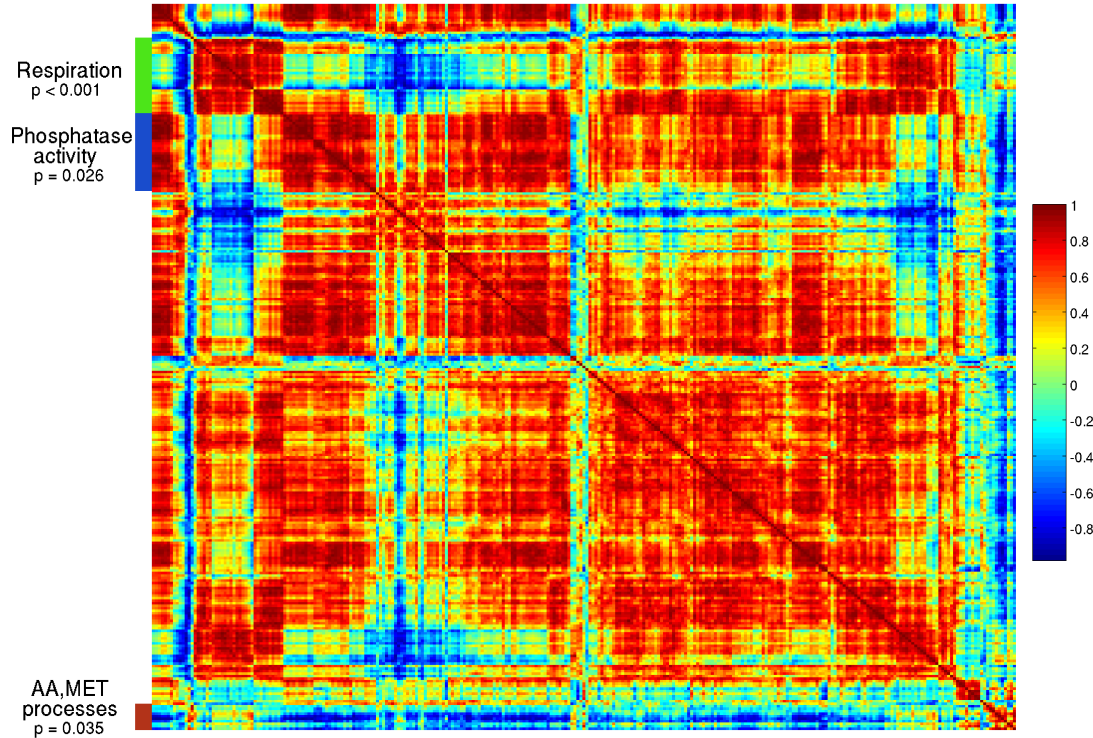
Figure 3.5: Correlation matrix for top beneficial mutants across all conditions. The correlation matrix represents the correlations between fitness profiles for 289 beneficial genetic perturbations across 12 conditions (6 environmental conditions for the deletion and overexpression). The resulting matrix was hierarchically clustered followed by GO enrichment.

3.2.3 Characterization of respiration genes related to the Warburg effect

We selected the cluster of genes (N = 30) enriched for respiration for further analysis because it has the highest average fitness effect in a single condition. The fitness profiles of these 30 mutants was hierarchically clustered and each resulting cluster was subject to enrichment analysis **(Figure 3.6).** We identified two significant biological modules: the pyruvate dehydrogenase complex (P < 0.001) and TORC signaling (P = 0.023). We also identified several genes related to mitochondria. The deletion of these 30 genes provides a beneficial advantage under amino acid limiting environments (SDG,

SDGR, and SDLGR), while having no significant effect in conditions supplemented with amino acids (SCG, SCCG, SCGR, SCCGR).

Pyruvate is metabolized via three major pathways: 1. Gluconeogenesis, 2. Ethanol fermentation, 3. Respiration **(Figure 3.7)**. Each pathway will generate a different amount of ATP, directly affecting the overall fitness of the cell. The pyruvate dehydrogenase complex (PDH) catalyzes the direct oxidative decarboxylation of pyruvate to acetyl-CoA, the precursor for the TCA cycle and oxidative phosphorylation. The increase in fitness is only detected when the PDH complex is disrupted on an amino acid limiting environment. One explanation is that the flux of pyruvate is forced down the fermentation pathway under amino acid starvation increasing immediate ATP production, while less control is exerted on the flux of pyruvate when not starved resulting in the metabolism of pyruvate via other pathways. In order to understand the molecular mechanism for this increase in fitness, we need to identify the flux of pyruvate in each pathway.

In the presence of glucose, *S. cerevisiae* exhibit an effect called glucose repression, where cells repress the expression of a larger number of genes that are required for the metabolism of alternate carbon sources including: respiratory, gluconeogenic, and galactose genes[63]. Under these conditions, the degree of respiration was measured to be zero, suggesting exclusive fermentation on glucose[64]. However, growth on galactose leads to simultaneous respiration and fermentation due to weaker repression of respiratory and gluconeogenic genes compared to glucose[64,65]. Respiration

activity was measured to be roughly half of a fully respiratory metabolism[64].

During growth on 2% galactose, the flux of pyruvate down the gluconeogenesis pathway should be minimal due to the repression of gluconeogenesis genes[65,66]. Deletion of phosphoenolpyruvate carboxykinase (PCK1), a required enzyme in gluconeogenesis, has no significant effect on fitness in SCG and SDG supporting this statement **(Figure 3.8)**. Therefore, the majority of pyruvate is processed via fermentation and respiration.

Under sufficient carbon availability yeast prefer fermentative metabolism, utilizing alcohol dehydrogenase (ADH1) to reduce acetylaldehyde to ethanol. Deletion of ADH1 decreases fitness by ~50%, most likely because cells are forced to produce ATP via respiration **(Figure 3.8)**. Respiration leads to increased ATP production compared to fermentation, however this process requires more proteins and the duration of a single cycle is significantly longer. Thus, respiration leads to lower growth rates but higher biomass yield in steady state conditions[64].

Deletion of genes in the PDH complex (PDA1, PDB1, LAT1) increases fitness by 30% only under amino acid limiting environments; no effect is detected in SCG **(Figure 3.8)**. This should increase the conversion of pyruvate to acetylaldehyde for two reasons: 1. Gluconeogenesis genes are downregulated, 2. Direct oxidative decarboxylation of pyruvate to acetyl-CoA via PDH is blocked. Acetylaldehyde can either be processed by alcohol dehydrogenase (ADH) yielding 2 ATP, NAD^+, and ethanol or by acetylaldehyde dehydrogenase (ALD) yielding acetate, which can then be

converted to acetyl-CoA via acetyl-CoA synthetase. The latter process is termed the pyruvate dehydrogenase bypass pathway, which occurs in the cytosol and yields acetyl-CoA by the action of the following enzymes: pyruvte decarboxylase (PDC), cytosolic acetalaldehyde dehydrogenase (ADH), and acetyl-CoA synthetase (ACS)[67]. This is an ATP-dependent process and thus reduces the net ATP yield, compared to the oxidative decarboxylation of pyruvate via PDH which does not require ATP hydrolysis. The majority of acetyl-CoA is synthesized via PDH because this is the more energetically favorable process. However it has been shown that PDH deficient strains maintain a fully respiratory metabolism through the pyruvate dehydrogenase bypass pathway[67].

The differential fitness increase in our PDH deficient strain (ΔPDA1) under SDG compared to SCG is most likely because the amino acid starvation response inhibits a member of the pyruvate dehydrogenase bypass pathway **(Figure 3.7)**. Therefore, when PDH is disrupted under SDG, pyruvate is mainly processed via fermentation increasing net ATP yield and thus growth rate. However, when PDH is disrupted under SCG, we believe that pyruvate is processed via fermentation and the pyruvate dehydrogenase bypass pathway yielding lower net ATP production and thus lower growth rate. This hypothesis is supported by the equal fitness effects between the ΔALD1 mutants under SCG and SDG **(Figure 3.8)**.
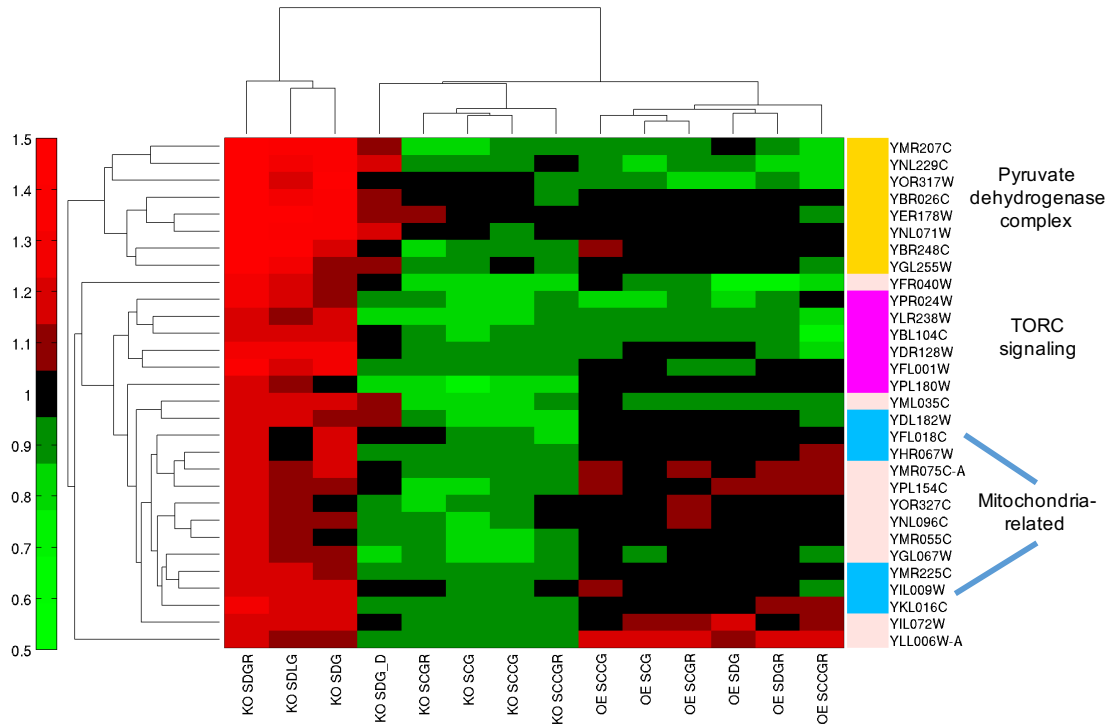
Figure 3.6: Clustering the fitness profiles of identified respiration genes reveals enrichment in the pyruvate dehydrogenase complex and TORC signaling. The identified respiration genes (N = 30) are hierarchically clustered, revealing significant enrichment for two biological modules: pyruvate dehydrogenase complex (P < 0.001) and TORC signaling (P = 0.023).
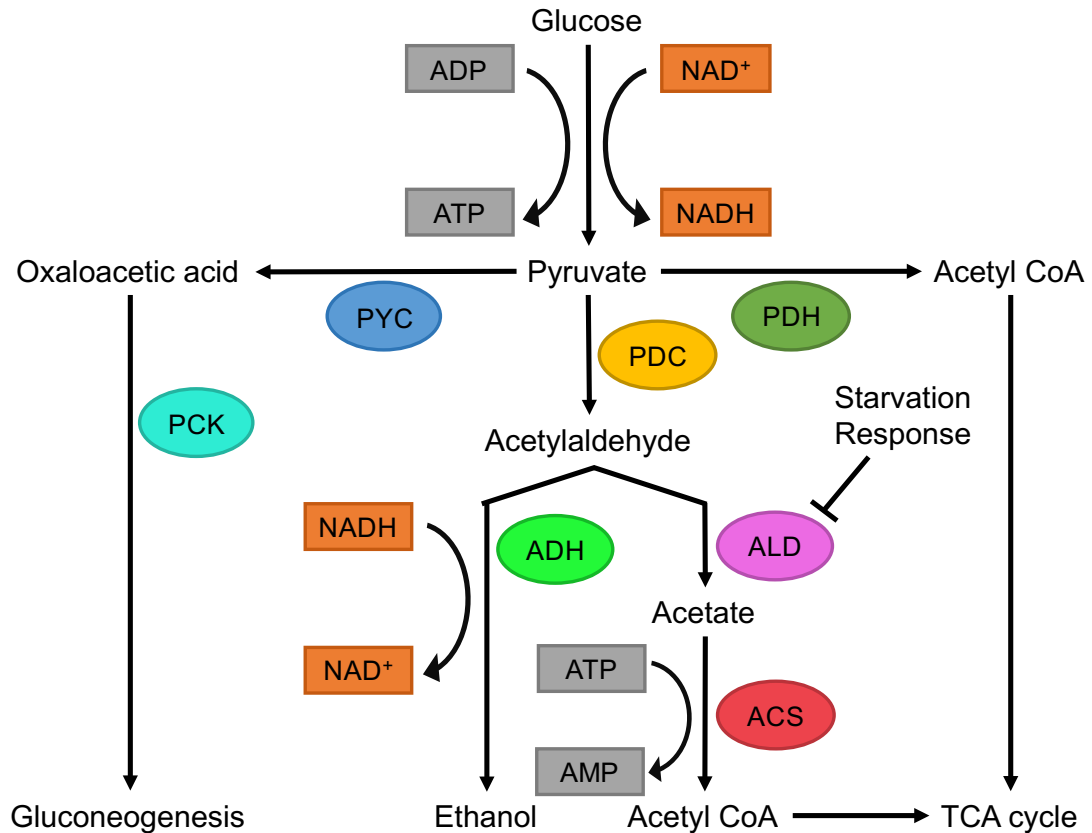
Figure 3.7: The cellular fate of pyruvate is determined by three major pathways: gluconeogenesis, fermentation, and respiration. Pyruvate is metabolized in each of three pathways via different enzymes: pyruvate carboxylase (PYC), pyruvate decarboxylase (PDC), and pyruvate dehydrogenase (PDH), respectively. The main products of these pathways are glucose-6-phosphate, ethanol/NAD^+/2 ATP, and NAD^+/32 ATP, respectively.
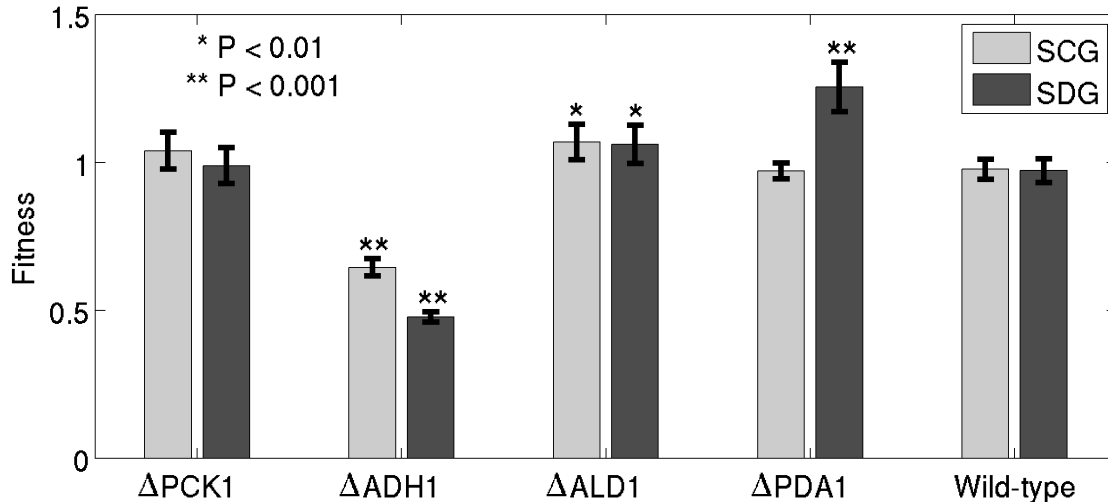
Figure 3.8: Deletion of key fermentative and respiratory enzymes reveals the cellular energy state. Deletion of PCK1, a key enzyme in gluconeogenesis, results in no significant change in fitness suggesting gluconeogenesis is not highly active. Deletion of ADH1, an enzyme required for the reduction of acetaldehyde to ethanol, results in significant decrease in fitness supporting high fermentative metabolism. Deletion of ALD1, an enzyme involved in the pyruvate dehydrogenase bypass pathway, results in slightly increased fitness suggesting small flux of pyruvate down this pathway. Deletion of PDA1, E1 alpha subunit of the pyruvate dehydrogenase complex (PDH), results in increased fitness only after SDG.

3.3 Discussion

We applied our framework to identify beneficial mutants across a range of metabolic conditions and different types of genetic perturbations. We found that positive fitness effects are highly dependent on the environment and type of genetic perturbations. There are similar numbers of deleterious strains between deletion and overexpression strains, however beneficial strains are significantly underrepresented in overexpression versus deletion strains. This is most likely because overexpression of gene products at ~1000x dosage is either neutral or deleterious.

Beneficial mutants may tell us more about the adaptive potential of the cell, or the ability to adapt to a novel environment. As mentioned previously, we now have two metrics to describe the adaptive potential of a cellular state: the fraction of beneficial mutants and the range of fitness effects of beneficial mutants. The fraction of beneficial mutants can be thought of as the proportion of the genome which when perturbed provides a beneficial effect, in essence the probability that a random perturbation will provide a beneficial effect. Therefore, the increase in this fraction could signify an increase in the adaptive potential of the cell. In some cases, the difference in fraction of beneficial mutants is also accompanied by an increase in the range of positive fitness effects **(Figure 3.3).** The range of positive fitness effects could also be related to the adaptive potential because perturbations with stronger positive effects fixate in a population in a shorter period of time, and thus increase the rate of adaptation. We calculated a larger fraction of beneficial mutants as well as a stronger positive fitness range in the amino acid starvation environments compared to environments supplemented with all amino acids. This could suggest that the adaptive potential under stress is higher than in the absence of stress because cells experience less selective pressure to survive in a neutral environment compared to an environmental stress.

We are also able to recover specific biological modules out of our identified beneficial mutants suggesting that at least a portion of beneficial mutations that occur in evolution are not random. These mutations provide a beneficial effect via a specific molecular mechanism related to the

environment, as evidenced by the increase in fitness from the deletion of respiration genes, specifically related to PDH activity, under galactose. Similar molecular phenotypes occur in cancer, where the repression of PDH activity is a key event known to provide a growth advantage in non-small cell lung and head and neck squamous cell carcinomas through its contribution to the Warburg effect[68,69]. *S. cerevisiae* has been discussed to be an effective system to study cancer due to similarity in energy metabolism, highlighted by a repression of oxidative metabolism and preference for aerobic glycolysis[56]. In our systematic screen, we were able to detect and quantify the fitness effect of loss of PDH activity in *S. cerevisiae*, which is known to have a similar metabolic state to cancer. The loss of genes involved in the PDH complex resulted in the largest fitness increase in our screen, which could signify the evolutionary advantage to repress these genes in cancer as opposed to the repression of other respiratory pathways.

Most of the time, genes representing the PDH complex are repressed in cancer instead of mutated resulting in non-functional PDH activity. This could be because it is not advantageous to lose PDH function for pleiotropic reasons. The tumor micro-environment is dynamic and in some cases the loss of PDH activity may be detrimental; for instance, the loss of PDH function in yeast under glucose-limited environments or environments with only non-fermentable carbon sources such as lactate results in significantly inhibited growth[66,67]. We were able to observe a similar pleiotropic effect in the ΔPDA1

mutant, where loss of PDH function resulted in a fitness increase in SDG but no effect in SCG.

The origin of the Warburg effect is not fully understood, however our current understanding is based on an increase in glucose uptake/glycolysis and/or down-regulation of mitochondrial metabolism. The repression of PDH function is a molecular mechanism known to contribute to this effect. However, our results could suggest that it is only advantageous to repress PDH function under amino acid starvation conditions. One possible theory for the origin of the Warburg effect is that cells in regions of tissue with low nutrient availability actively downregulate PDH function in order to survive.

Our methods may yield insights into cancer evolution and the mechanisms it utilizes to gain growth advantages under specific environments. Therefore, screening multiple environments for beneficial mutants may paint a better picture of the cellular states necessary for specific beneficial events to manifest. The fact that we are able to quantify the effect of a molecular phenotype known to provide a beneficial effect in cancer is evidence that we can use this method to quantify the effects of other mechanisms in cancer. We may then be able to determine which mechanism provides the strongest beneficial effect in cancer, resulting in potentially better drug targets.

Chapter 3, in part is currently being prepared for submission for publication of the material. Hsu, Brian; Medetgul-Ernar, Kate; Hines, Cameron; Michaca, Manuel; Regent, Nick; Carvunis, Anne; Ideker, Trey. The thesis author was the primary investigator and author of this paper.

Chapter 4. Environmental history affects evolutionary outcome

4.1 Introduction

Organisms are constantly exposed to a wide range of environments. In order to survive they must adapt to these environmental changes. This can be achieved through a variety of methods including regulation of gene expression[2] and/or mutations. The fundamental question is what contributes to this rate of adaptation and whether or not populations will be able to adapt rapidly enough to avoid extinction. Classic evolutionary models estimate the adaptation rate based on the allelic variation and Mendelian inheritance[70]. However, there has been recent evidence suggesting that environmentally generated variation can also influence the adaptation rate as well. Empirical studies have shown that these environmentally generated effects are transmitted across generations via nongenetic inheritance mechanisms. Nongenetic inheritance includes transmission of epigenetic variation (i.e. DNA-methylation patterns, RNA), parental glandular secretions, nutrients, hormones, and behaviors to offspring[71]. These patterns have been previously observed in transgenerational studies[72,73].

To further understand the mechanisms of epigenetic inheritance, other studies have performed systematic analysis of *S. cerevisiae* in pretreatment experiments. They found that yeast use information from their environment to prepare for future threats, a phenomenon known as acquired stress reistance[74]. This allows cells exposed to a mild dose of one stress to survive

an otherwise lethal dose of a second stress and has been observed in diverse organisms ranging from bacteria to humans[3,75,]. These studies identify genes involved in acquired stress resistance by looking for slow growth/lethal phenotypes in strains following pretreatment. Thus, we have been able to thoroughly characterize the transcriptional program involved in the stress response in *S. cerevisiae*, also known as the environmental stress response (ESR)[1].

Although we know the past environment experienced by the cell affects its growth on the subsequent environment, it is still unknown whether this effect can cause cells to grow better. Therefore, we systematically tested whether the environmental history experienced by the cell could provide a beneficial effect for growth in subsequent conditions. We pretreated ~8000 engineered *Saccharomyces cerevisiae* deletion and overexpression mutants on either an amino acid limiting environment or an environment supplemented with all amino acids and then grew all mutants on 6 different environmental conditions. We then compared the fitness on each of the 6 conditions after growth on the pretreatment conditions to determine whether the environmental history can selectively provide a beneficial effect.

4.2 Results

4.2.1 Identification of past-dependent mutants.

We estimated fitness in each of the 6 environmental conditions listed in **Table 2** after either an amino acid limiting environment or an environment supplemented with all amino acids. We calculated the fitness differential for

each mutant (N = 4) by taking the difference between fitness estimates after each pretreatment condition. We applied the same calculation to the WT strain (N = 3072) to estimate the null distribution. We then compared the fitness differential of each strain to the null distribution using a Mann Whitney U test to determine if there is a significant difference between the underlying distributions, followed by multiple testing hypothesis correction using the Q-value. The threshold used to identify significant hits was a fitness differential above two standard deviations of the null distribution **(Figure 4.1).** We named these our "past-dependent" mutants. To identify past-dependent beneficial mutants we took our set of past-dependent mutants and and filtered out candidates that were beneficial only after one pretreatment condition but not the other **(Figure 4.3A)**.

We correlated fitness estimations on SCCG after different pretreatment conditions and after the same pretreatment condition and found that fitness estimations are highly correlated (r = 0.97) after the same pretreatment condition and less strongly correlated (r = 0.82) after different pretreatment conditions. This supports that the environmental history does affect fitness on subsequent environments **(Figure 4.2).**

To check the accuracy of our pipeline in identifying past-dependent beneficial mutants, we applied this analysis to mutants growing on SCCG after the same pretreatment condition and different pretreatment conditions. The analysis of mutants growing on SCCG after the same pretreatment conditions served as a negative control. We identify 48 past-dependent beneficial

mutants, 43 after SCG and 5 after SDG **(Figure 4.3B).** In our negative control, we detect 0 past-dependent beneficial mutants.

We also applied this analysis to the SDGR condition and identified 60 past-dependent mutants, 53 after SCG and 7 after SDG (**Figure 4.3B**). In both of the conditions tested, we find significantly more past-dependent beneficial mutants after pretreatment in an environment supplemented with all amino acids compared to an amino acid limiting environment.
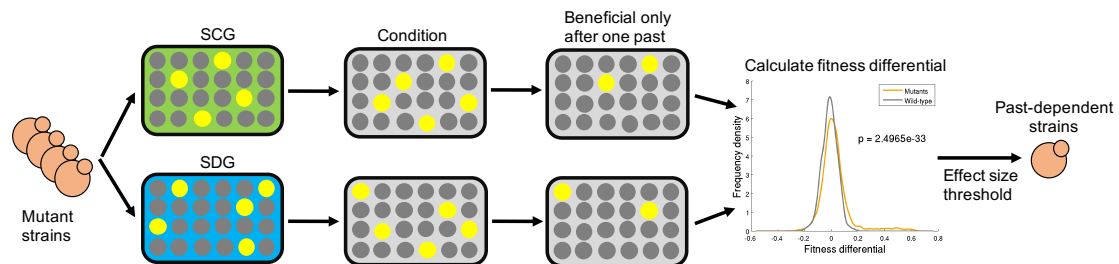
Figure 4.1: Experimental pipeline to identify past-dependent beneficial mutants. Deletion and overexpression mutants are pretreated on either an environment supplemented with all amino acids (SCG) or an amino acid limiting environment (SDG) and subsequently grown on a new environmental condition. Beneficial mutants are identified in each condition based on methods described previously. The fitness differential is calculated by calculating the difference between fitness effects on a condition after each past. The null distribution is estimated by calculating the fitness differential between the wild-type strains. Beneficial past-dependent mutants are identified as beneficial only after one past and also having significant fitness differential.
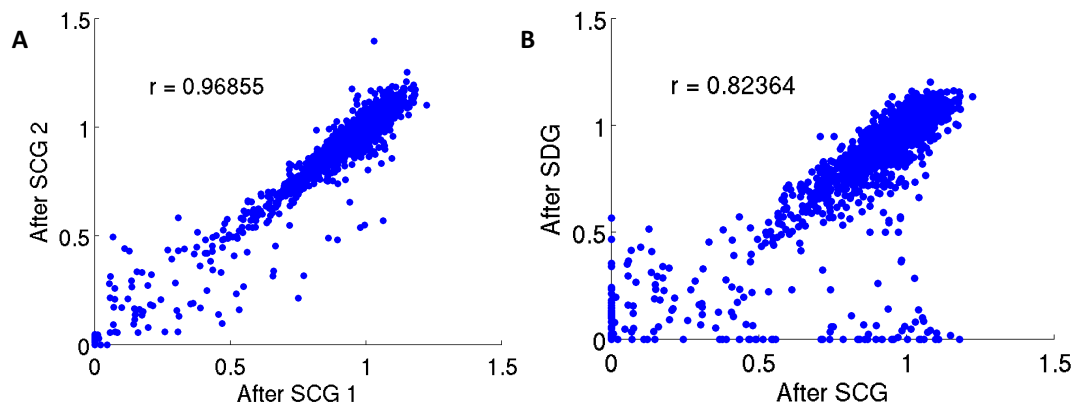
Figure 4.2: Fitness estimations are less correlated after different past environments. (A) Fitness estimations on SCCG after growth on SCG are highly correlated (same past environment) (r = 0.97). **(B)** Fitness estimations on SCCG after growth on SCG or SDG are less correlated (different past environment) (r = 0.82).
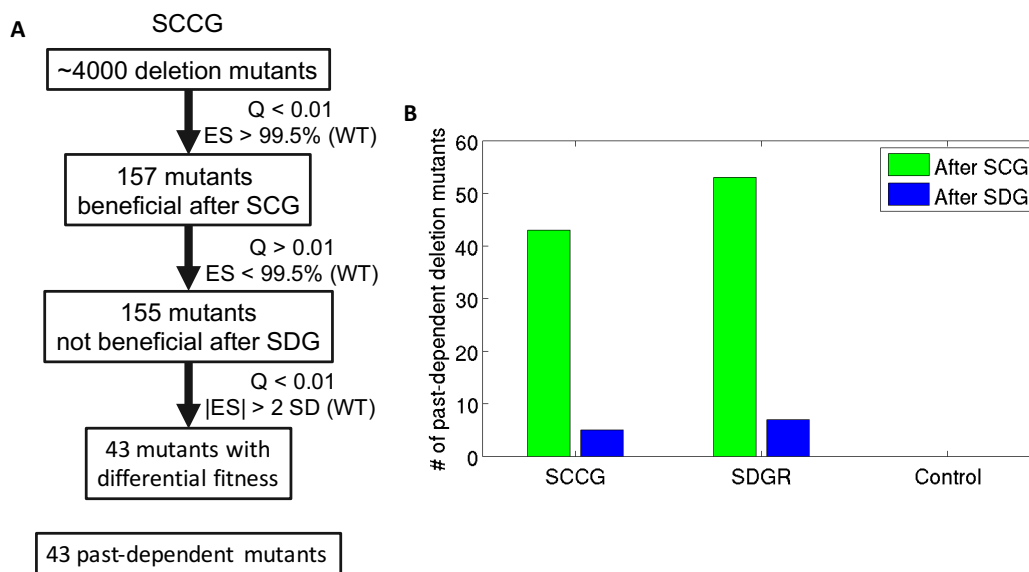
Figure 4.3: Identification of past-dependent beneficial deletion mutants. For a beneficial mutant to be considered past-dependent, it must have a fitness differential >97.5% or <2.5% of the null distribution (fitness differential of the WT strain) and be beneficial only after one past. **(A)** Beneficial and deleterious mutants are classified as past-dependent in a positive (different pasts) and negative (same pasts) control experiment. **(B)** Past-dependence is determined for beneficial mutants across six environmental conditions.

4.2.2 Validation of past-dependent beneficial mutants.

We tested all of our past-dependent beneficial mutants in SCCG and SDGR in an independent low-throughput 1536-density colony assay experiment. We applied a similar pipeline as before, however now we only use an effect size threshold of two standard deviations **(Figure 4.4)**. We validated 2 past-dependent mutants in SCCG and 3 past-dependent mutants in SDGR. This low validation rate could be due to a few reasons: 1. Increase in variance of the wild-type distribution, 2. Inherent differences in experimental methodologies. As mentioned previously, the variance of the WT distribution in the validation experiment is larger than that of the genome-wide screen. This causes the effect size threshold used previously to significantly underestimate the same effect size threshold in the validation experiment, i.e. two standard deviations of the old WT distribution is significantly smaller than two standard deviations of the new WT distribution. Therefore, we pick up more false positives due to this difference in the underlying WT distribution. In higher density plates, individual colonies are also closer together and may exhibit competitive effects in terms of nutrient availability and space limitations. This may lead to cross-colony spatial interactions, which are non-existent at lower-density formats. However, the 2-3 past-dependent beneficial mutants we validated in each condition have the highest effect size and fitness differential in the genome-wide screen suggesting that our problem lies in the effect size threshold.

SCCG

SDGR

| 40 past-dependent mutants |

ES > 2 SD (WT)

| 20 mutants beneficial after SCG |

ES < 2 SD (WT)

| 13 mutants not beneficial after SDG |

|ES| > 2 SD (WT)

| 2 mutants with differential fitness |

| 2 validated past-dependent mutants |

| 53 past-dependent mutants |

ES > 2 SD (WT)

| 3 mutants beneficial after SCG |

ES < 2 SD (WT)

| 3 mutants not beneficial after SDG |

|ES| > 2 SD (WT)

| 3 mutants with differential fitness |

| 3 validated past-dependent mutants |

Figure 4.4: Validation of past-dependent beneficial mutants. We randomly selected 40 and 53 beneficial past-dependent mutants in SCCG and SDGR, respectively, for a follow-up validation experiment. We applied a similar pipeline as in **(Figure 4.3A)**, however this time we only used an effect size threshold of two standard deviations. We validated 2 and 3 beneficial past-dependent mutants in SCCG and SDGR, respectively.

4.2.3 Characterization of past-dependent beneficial mutants.

To determine whether the fitness effects of past-dependent beneficial mutants are transient, we grew each mutant on SCCG and SDGR for 8 days, the equivalent of 4 pin-transfers. For the three beneficial past-dependent mutants (YDR128W, YFL001W, YK016C) on SDGR, we see that initially on pin 1 there is a significant differential fitness effect. However, this effect diminishes over time and completely disappears during pin 2 **(Figure 4.5)**.

This suggests that any increase in fitness is only temporary and after sustained growth in the novel environment the memory of the environmental history will eventually disappear, as shown by the decrease in differential fitness from Pin 2 to Pin 4 **(Figure 4.5)**.

To understand the expected behavior of beneficial mutants, we plotted the fitness over time of 9 beneficial past-independent mutants (beneficial after either SCG or SDG) in SDGR **(Figure 4.6)**. In each of the 9 beneficial past-independent mutants, the fitness differential is almost non-existent on all 4 pins. There is a significant difference between the fitness patterns for the beneficial past-dependent and past-independent mutants. The fitness of the beneficial past-dependent mutants also fluctuates more than their counterparts, which could suggest an important role in regulation. YDR138W (MTC5) is a Seh1-associated complex which dynamically associates with the vacuolar membrane, regulates TORC1 signaling, and is involved in intracellular trafficking, amino acid biogenesis, and response to nitrogen starvation[76]. YFL001W (DEG1) is a tRNA pseudouridine synthase which introduces pseudouridines at position 38 or 39 in tRNA[77]. YKL016C (ATP7) encodes for subunit d of the stator stalk of mitochondrial F1F0 ATP synthase. Based on their functions, MTC5 and DEG1 may be involved with the mechanism to sense the current nutrient and translational capacity, respectively, of the cell. Therefore, disruption of these genes would lead to the fluctuations in fitness we observe. Initially, ATP7 seems unrelated to the other

two genes as being involved in regulation. However, this could suggest a new regulatory role for ATP7.

Out of the 9 beneficial past-dependent mutants, three of them (YER178W, YMR207C, YNL071W) interestingly have similar fitness patterns over time that are different than the other six **(Figure 4.6)**. These mutants were identified previously as part of the pyruvate dehydrogenase complex in Chapter 2. This could provide evidence that similar molecular phenotypes result in similar fitness estimates. The fitness of each of these three mutants also increases over time from pin 1 to pin 4 whereas the fitness of the other six remains constant in the same period of time, which could mean they are involved in a molecular mechanism with a temporal effect.
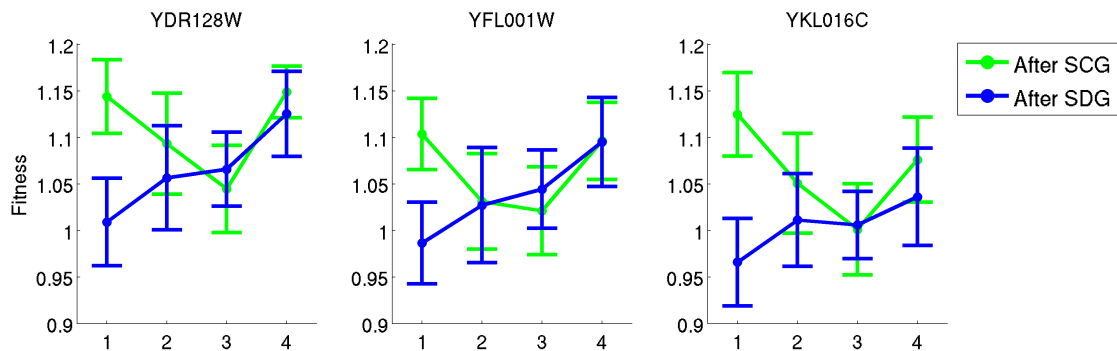
Figure 4.5: Dynamics of past-dependent beneficial mutants. The fitness estimates of the 3 validated beneficial past-dependent mutants on SDGR over 4 pin-transfers after either SCG or SDG are plotted. The differential fitness observed during pin 1 decreases over time and disappears after pin 2. All 3 mutants appear to approach the same steady-state fitness value in SDGR.
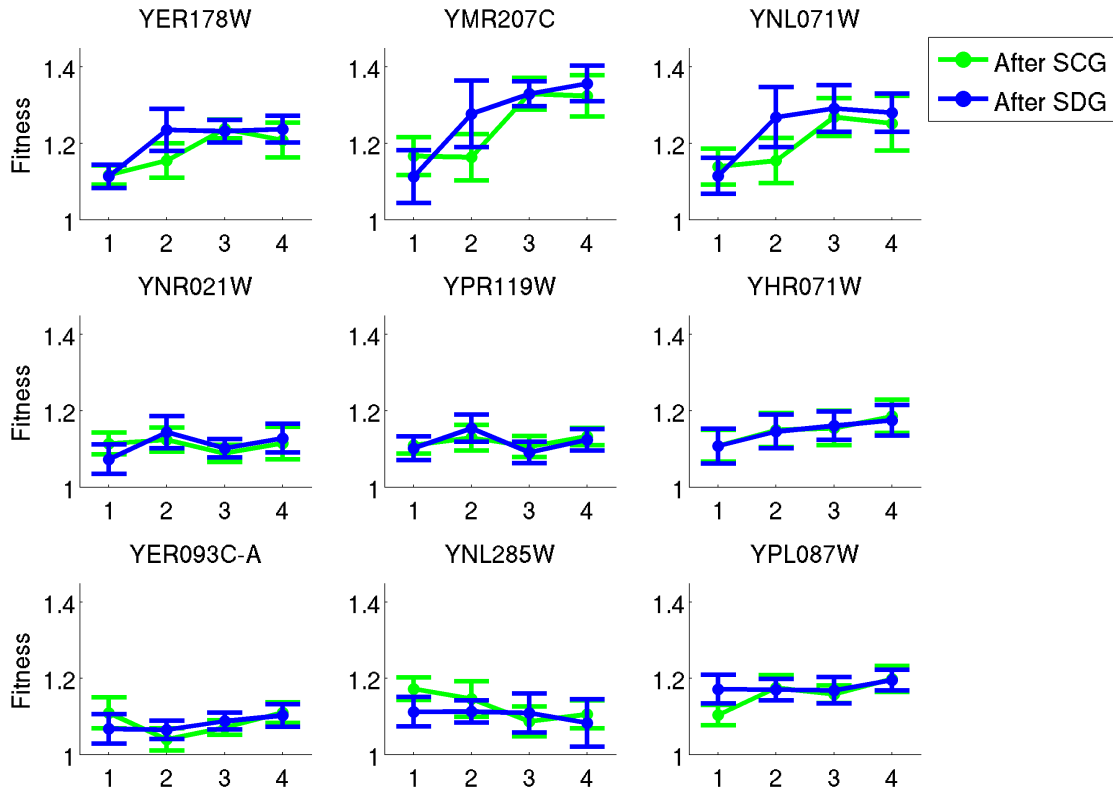
Figure 4.6: Dynamics of past-independent beneficial mutants. The fitness estimates of 9 validated beneficial past-independent mutants on SDGR over 4 pin-transfers after either SCG or SDG are plotted. The fitness differentials of these past-independent mutants is almost non-existent across all 4 pins. YER178W, YMR207C, YNL071W all exhibit similar dynamic fitness profiles; this set of genes is involved in the pyruvate dehydrogenase complex.

4.3 Discussion

We systematically tested whether the environmental history experienced by the cell can provide a beneficial effect to the current cellular state. We found that this phenomenon occurs in certain mutants, which have a memory of their environmental history and thus provides a growth advantage in the subsequent condition. This is supported by the differential fitness effects for a single mutant following pretreatment on either SCG or SDG. However, this effect is less pervasive and is only seen in <1% of beneficial mutants.

One possible explanation for this phenomenon could be the temporal aspect of transcriptional reprogramming that occurs in a novel environment. Each environment has an associated optimal transcriptional state, or the level of expression of each gene which will result in ideal cellular growth. When cells transition to a novel environment they must reprogram their transcriptome, which was optimized for growth in the old environment. The deletion of a gene could potentially alter the reprogramming process. If a gene product required for growth in an environment is missing, the cell will have to compensate for this loss. This could be through the upregulation of another pathway or the activation of a dormant pathway. However, the compensation for one pathway could either be sufficient or cause a chain-effect requiring modification of other pathways. The first option would result in a shorter time for reprogramming, while the second option would result in a longer period for reprogramming. Another possibility is that the gene product that is lost is a key regulator gene, which controls multiple pathways. Therefore, when this gene is deleted the cell must compensate for all of the processes/functions which that gene was controlled. Thus, increasing the amount of pathways which must be compensated will likely increase the amount of time for transcriptional reprogramming to occur. The second option seems more probable and could explain the temporal effects of fitness over time for the beneficial past-dependent genes.

This phenomenon is not apparent for the majority of mutants and thus these cells are able to reprogram their transcriptome with ease. However, for

2-3 beneficial past-dependent mutants they're fitness significantly fluctuates, which could signify the active effort of the cell to reach the optimal transcriptional state.

Our results show that before pin 2, beneficial past-dependent mutants will have a growth advantage depending on their environmental history. Therefore, there is a period of time when a strain with the same genotype will have a selective advantage other strains of the same genotype after a specific environment. During this period, a single strain could potentially take over the population or an adaptive event may occur. This period is only around ~2 days for *S. cerevisiae*. However, for other organisms with longer generation times this effect may last for days, months, or even years. During this longer time span, several changes could occur such as a mutation which could change the evolutionary outcome between two identical strains with the same genotype but different environmental histories.

Chapter 4, in part is currently being prepared for submission for publication of the material. Hsu, Brian; Medetgul-Ernar, Kate; Hines, Cameron; Michaca, Manuel; Regent, Nick; Carvunis, Anne; Ideker, Trey. The thesis author was the primary investigator and author of this paper.

Chapter 5: Discussion

Beneficial mutations are paramount in the process of evolution. However, the quantitative nature of these events remains elusive. There have been very few studies regarding beneficial mutations, especially quantifying their effects and how they contribute to the rate of adaptation. We developed a method to study beneficial mutants and found that the fraction of beneficial mutants and range of fitness effects is dependent on the environmental condition. These two metrics could potentially be used to estimate the rate of adaptation and positive selective pressure. This could be used to quantify the rate of adaptation for cancer cells in different environmental conditions to determine if a specific environment is causing cancer cells to evolve at a faster rate. These metrics could also be compared to currently estimated rates of beneficial mutations to understand whether they agree and can recapitulate the same information. If in agreement, our method would provide a significantly easier alternative to estimate the rate of beneficial mutations compared to the traditional and costly method via E&R experiments.

We found that the deletion of genes encoding the pyruvate dehydrogenase complex, a molecular phenotype also observed in cancer, causes the largest increase in fitness. By quantifying the fitness effects, we may be able to calculate the significance of each molecular mechanism to the overall survival of the cancer cell. Thus, if we can detect the molecular changes that will be selected for and occur from an evolutionary perspective in

cancer, we may be able to gain a better understanding of cancer progression and potentially slow down the rate of development.

Ultimately, we have developed a platform to study beneficial mutations and adaptive evolution. For the past few decades, empirical research in this field has lagged due to the lack of technological tools. However, we hope to provide the necessary tools for others to utilize and advance the current understanding in this field. Deepening our understanding of beneficial mutations will definitely lead to great strides in understanding the evolution of cancer.

Chapter 6: Methods

Yeast strains and media

All deletion mutant strains used are based on the commercially available yeast knockout (YKO) strain collection (Thermo Fisher Scientific Inc., Waltham/MA) with kanamycin as a deletion marker, the construction which has been described previously[13]. The YKO collection contains deletion of a nonessential gene per haploid strain (BY4741). The overexpression mutants strains used are based on the barFLEX collection[36]. All overexpression mutants were grown on the same exact conditions as the deletion strains except URA was dropped out in each media. Yeast strains were grown in various medias listed in **Table 2**. Amino acid dropout mix was made as described previously[78].

Genome-wide screen

Each genome-wide screen started from a library of mutants maintained on a set of 1536-density plates. This set of plates is pinned on to either SCG or SDG for two consecutive pins. On the 3rd pinning, the plates are pinned on to one of the medias listed in **Table 2**.

Fitness estimation

Agar-to-agar yeast transfers were conducted using a Singer RoToR robotic plate handler (Singer Instrument Co. Ltd). Each agar plate was imaged at multiple time points using a commercially available SLR camera (18Mpixel Rebel T3i, Canon USA Inc., Melville/ NY) with an 18–55 mm zoom lens.

Images were normalized, spatially corrected, and quantified using a set of custom algorithms (aka "The Colony Analyzer Toolkit") written in Matlab (MathWorks Inc., Natick/MA)[17,32].

Identification of beneficial genetic perturbations

We used the ΔHO (YDL227C) mutant as the WT reference strain to estimate the relative fitness of every deletion strain in the genome-wide screen. We calculated the significance between each deletion strain and the WT reference using the Mann-Whitney U test. To correct for multiple testing hypothesis, each p-value was converted to a Q-value[79]. Each deletion strain was classified as beneficial, deleterious, and neutral based on a Q-value threshold (Q < 0.01) and a fitness threshold (5% and 95% of WT).

Calculation of reproducibility rate

The reproducibility rate of beneficial mutants between two independent experiments were calculated as follows: subsample the mutants to include only those with a fitness greater than the median of the WT fitness for each experiment, randomly sample the number of identified beneficial mutants from each subsampled set, find the overlap between the two sets, find the unique number of mutants from the union of both sets, divide the number of overlapping mutants by the number of unique mutants.

Hierarchical clustering

Hierarchical clustering analysis was performed in Matlab using the clustergram function. Values in each column were standardized to reduce

noise between experiments. Distance metric used is Euclidean distance and the distance between clusters was calculated using the average distance.

Validation of beneficial genetic perturbations

Validation experiments were performed in a 1536-density colony array with N = 40 replicates per mutant. 288 mutants were randomly selected for follow up validation. 20 control strains were used to estimate the wild-type distribution: 14 randomly selected mutants from the barFLEX collection[36] with their plasmids removed following 5-fluoroorotic acid (5-FOA) treatment, 5 pseudogene strains (YLL016W, YLL017W, YFL056C YCL075W, YIL170W), and the ΔHO (YDL227C). All mutant and control strains were randomly arrayed onto a single plate using the Tecan Freedom EVO100. All mutants were grown on either SCG or SDG for two consecutive pin-transfers. All mutants were then pinned on to SCCG and SDGR and grown for 4 pin-transfers over a span of 8 days. Fitness was estimated every 48 hours. Mutants in the experiment were considered as validated if their fitness was beyond larger than two standard deviations than the wild-type distribution.

Chapter 7: Bibliography

1. Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. & Brown, P. O. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol. Biol. Cell* **11,** 4241–4257 (2000).

2. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9,** 583–593 (2008).

3. Berry, D. B. & Gasch, A. P. Stress-activated Genomic Expression Changes Serve a Preparative Role for Impending Stress in Yeast. *Mol. Biol. Cell* **19,** 4580–4587 (2008).

4. Darwin, C. *On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life*. (1859).

5. Haldane, J. B. S. *The causes of evolution*. (Longmans, Green and Co., 1932).

6. Neu, H. C. The Crisis in Antibiotic Resistance. *Science* **257,** 1064–1073 (1992).

7. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).

8. Araya, C. L., Payen, C., Dunham, M. J. & Fields, S. Whole-genome sequencing of a laboratory-evolved yeast strain. *BMC Genomics* **11,** 88 (2010).

9. Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D. & Desai, M. M. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500,** 571–574 (2013).

10. Caspeta, L. & Nielsen, J. Thermotolerant Yeast Strains Adapted by Laboratory Evolution Show Trade-Off at Ancestral Temperatures and Preadaptation to Other Stresses. *mBio* **6,** e00431-15 (2015).

11. Payen, C., Sunshine, A. B., Ong, G. T., Pogachar, J. L., Zhao, W. & Dunham, M. J. Empirical determinants of adaptive mutations in yeast experimental evolution. *bioRxiv* 14068 (2015). doi:10.1101/014068

12. Carpenter, A. E. & Sabatini, D. M. Systematic genome-wide screens of gene function. *Nat. Rev. Genet.* **5,** 11–22 (2004).

13. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kötter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W. & Johnston, M. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418,** 387–391 (2002).

14. Giaever, G. & Nislow, C. The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics* **197,** 451–465 (2014).

15. Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y. & Wei, W. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509,** 487–491 (2014).

16. Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G. & Zhang, F. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343,** 84–87 (2014).

17. Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., Hibbs, M., Hess, D., Gingras, A.-C., Bader, G. D., Troyanskaya, O. G., Brown, G. W., Andrews, B., Boone, C. & Myers, C. L. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7,** 1017–1024 (2010).

18. Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D. F., Chu, A. M., Jordan, M. I., Arkin, A. P. & Davis, R. W. Chemogenomic profiling: Identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 793–798 (2004).

19. Li, Z., Vizeacoumar, F. J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F. S., Min, R., VanderSluis, B., Bellay, J., DeVit, M., Fleming, J. A., Stephens,

A., Haase, J., Lin, Z.-Y., Baryshnikova, A., Lu, H., Yan, Z., Jin, K., Barker, S., Datti, A., Giaever, G., Nislow, C., Bulawa, C., Myers, C. L., Costanzo, M., Gingras, A.-C., Zhang, Z., Blomberg, A., Bloom, K., Andrews, B. & Boone, C. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat. Biotechnol.* **29,** 361–367 (2011).

20. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar–coding strategy. *Nat. Genet.* **14,** 450–456 (1996).

21. Pierce, S. E., Davis, R. W., Nislow, C. & Giaever, G. Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. *Nat. Protoc.* **2,** 2958–2974 (2007).

22. Memarian, N., Jessulat, M., Alirezaie, J., Mir-Rashed, N., Xu, J., Zareie, M., Smith, M. & Golshani, A. Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC Bioinformatics* **8,** 117 (2007).

23. Delneri, D., Hoyle, D. C., Gkargkas, K., Cross, E. J. M., Rash, B., Zeef, L., Leong, H.-S., Davey, H. M., Hayes, A., Kell, D. B., Griffith, G. W. & Oliver, S. G. Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat. Genet.* **40,** 113–117 (2008).

24. Sliwa, P. & Korona, R. Loss of dispensable genes is not adaptive in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17670–17674 (2005).

25. Lang, G. I. & Murray, A. W. Estimating the Per-Base-Pair Mutation Rate in the Yeast Saccharomyces cerevisiae. *Genetics* **178,** 67–82 (2008).

26. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. A. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci.* **111,** E2310–E2318 (2014).

27. Teng, X., Dayhoff-Brannigan, M., Cheng, W.-C., Gilbert, C. E., Sing, C. N., Diny, N. L., Wheelan, S. J., Dunham, M. J., Boeke, J. D., Pineda, F. J. & Hardwick, J. M. Genome-wide Consequences of Deleting Any Single Gene. *Mol. Cell* **52,** 485–494 (2013).

28. Celiker, H. & Gore, J. Competition between species can stabilize public-goods cooperation within a species. *Mol. Syst. Biol.* **8,** 621 (2012).

29. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc. Natl. Acad. Sci.* **104,** 1877–1882 (2007).

30. Steinmetz, L. M., Scharfe, C., Deutschbauer, A. M., Mokranjac, D., Herman, Z. S., Jones, T., Chu, A. M., Giaever, G., Prokisch, H., Oefner, P. J. & Davis, R. W. Systematic screen for human disease genes in yeast. *Nat. Genet.* **31,** 400–404 (2002).

31. Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St.Onge, R. P., Tyers, M., Koller, D., Altman, R. B., Davis, R. W., Nislow, C. & Giaever, G. The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. *Science* **320,** 362–365 (2008).

32. Bean, G. J., Jaeger, P. A., Bahr, S. & Ideker, T. Development of Ultra-High-Density Screening Tools for Microbial 'Omics'. *PLOS ONE* **9,** e85177 (2014).

33. Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The Genomic Landscape and Evolutionary Resolution of Antagonistic Pleiotropy in Yeast. *Cell Rep.* **2,** 1399–1410 (2012).

34. Gerstein, A. C., Cleathero, L. A., Mandegar, M. A. & Otto, S. P. Haploids adapt faster than diploids across a range of environments. *J. Evol. Biol.* **24,** 531–540 (2011).

35. Mable, B. K. Ploidy evolution in the yeast Saccharomyces cerevisiae: a test of the nutrient limitation hypothesis. *J. Evol. Biol.* **14,** 157–170 (2001).

36. Douglas, A. C., Smith, A. M., Sharifpoor, S., Yan, Z., Durbic, T., Heisler, L. E., Lee, A. Y., Ryan, O., Göttert, H., Surendra, A., van Dyk, D., Giaever, G., Boone, C., Nislow, C. & Andrews, B. J. Functional Analysis With a Barcoder Yeast Gene Overexpression System. *G3 GenesGenomesGenetics* **2,** 1279–1289 (2012).

37. Li, W.-H., Gojobori, T. & Nei, M. Pseudogenes as a paradigm of neutral evolution. *Nature* **292,** 237–239 (1981).

38. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, 1983).

39. Orr, H. A. The population genetics of beneficial mutations. *Philos. Trans. R. Soc. B Biol. Sci.* **365,** 1195–1201 (2010).

40. Jerby-Arnon, L., Pfetzer, N., Waldman, Y. Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P. A., Gottlieb, E. & Ruppin, E. Predicting Cancer-Specific Vulnerability via Data-Driven Detection of Synthetic Lethality. *Cell* **158,** 1199–1209 (2014).

41. Srivas, R., Shen, J. P., Yang, C. C., Sun, S. M., Li, J., Gross, A. M., Jensen, J., Licon, K., Bojorquez-Gomez, A., Klepper, K., Huang, J., Pekin, D., Xu, J. L., Yeerna, H., Sivaganesh, V., Kollenstart, L., van Attikum, H., Aza-Blanc, P., Sobol, R. W. & Ideker, T. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. *Mol. Cell* **63,** 514–525 (2016).

42. Hofree, M., Shen, J. P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat. Methods* **10,** 1108–1115 (2013).

43. The Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. & Stuart, J. M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45,** 1113–1120 (2013).

44. Williams, G. C. Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution* **11,** 398–411 (1957).

45. He, X. & Zhang, J. Toward a Molecular Understanding of Pleiotropy. *Genetics* **173,** 1885–1891 (2006).

46. Bowman, J. C. Genotype × environment interactions. *Ann. Génétique Sélection Anim.* **4,** 117 (1972).

47. Gagneur, J., Stegle, O., Zhu, C., Jakob, P., Tekkedil, M. M., Aiyar, R. S., Schuon, A.-K., Pe'er, D. & Steinmetz, L. M. Genotype-Environment Interactions Reveal Causal Pathways That Mediate Genetic Effects on Phenotype. *PLOS Genet* **9,** e1003803 (2013).

48. Oijen, M. G. C. T. van & Slootweg, P. J. Gain-of-Function Mutations in the Tumor Suppressor Gene p53. *Clin. Cancer Res.* **6,** 2138–2145 (2000).

49. Zhou, G., Wang, J., Zhao, M., Xie, T.-X., Tanaka, N., Sano, D., Patel, A. A., Ward, A. M., Sandulache, V. C., Jasser, S. A., Skinner, H. D., Fitzgerald, A. L., Osman, A. A., Wei, Y., Xia, X., Songyang, Z., Mills, G. B., Hung, M.-C., Caulin, C., Liang, J. & Myers, J. N. Gain-of-Function Mutant p53 Promotes Cell Growth and Cancer Cell Metabolism via Inhibition of AMPK Activation. *Mol. Cell* **54,** 960–974 (2014).

50. Fabbro, M., Savage, K., Hobson, K., Deans, A. J., Powell, S. N., McArthur, G. A. & Khanna, K. K. BRCA1-BARD1 Complexes Are Required for p53Ser-15 Phosphorylation and a G1/S Arrest following Ionizing Radiation-induced DNA Damage. *J. Biol. Chem.* **279,** 31251–31258 (2004).

51. Kim, H., Chen, J. & Yu, X. Ubiquitin-Binding Protein RAP80 Mediates BRCA1-Dependent DNA Damage Response. *Science* **316,** 1202–1205 (2007).

52. Wang, B., Matsuoka, S., Ballif, B. A., Zhang, D., Smogorzewska, A., Gygi, S. P. & Elledge, S. J. Abraxas and RAP80 Form a BRCA1 Protein Complex Required for the DNA Damage Response. *Science* **316,** 1194–1198 (2007).

53. Kiers, J., Zeeman, A.-M., Luttik, M., Thiele, C., Castrillo, J. I., Steensma, H. Y., Van Dijken, J. P. & Pronk, J. T. Regulation of alcoholic fermentation in batch and chemostat cultures of Kluyveromyces lactis CBS 2359. *Yeast* **14,** 459–469 (1998).

54. DE DEKEN, R. H. The Crabtree Effect: A Regulatory System in Yeast. *Microbiology* **44,** 149–156 (1966).

55. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324,** 1029–1033 (2009).

56. Diaz-Ruiz, R., Rigoulet, M. & Devin, A. The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochim. Biophys. Acta BBA - Bioenerg.* **1807,** 568–576 (2011).

57. Adams, B. G. Induction of Galactokinase in Saccharomyces cerevisiae: Kinetics of Induction and Glucose Effects. *J. Bacteriol.* **111,** 308–315 (1972).

58. Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S. G., Cyert, M., Hughes, T. R., Boone, C. & Andrews, B. Mapping Pathways and Phenotypes by Systematic Gene Overexpression. *Mol. Cell* **21,** 319–330 (2006).

59. Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G. & Marton, M. J. Transcriptional Profiling Shows that Gcn4p Is a Master Regulator of Gene Expression during Amino Acid Starvation in Yeast. *Mol. Cell. Biol.* **21,** 4347–4368 (2001).

60. Hinnebusch, A. G. Translational Regulation of Gcn4 and the General Amino Acid Control of Yeast. *Annu. Rev. Microbiol.* **59,** 407–450 (2005).

61. Hillenmeyer, M. E., Ericson, E., Davis, R. W., Nislow, C., Koller, D. & Giaever, G. Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action. *Genome Biol.* **11,** R30 (2010).

62. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

63. Gancedo, J. M. Yeast Carbon Catabolite Repression. *Microbiol. Mol. Biol. Rev.* **62,** 334–361 (1998).

64. Fendt, S.-M. & Sauer, U. Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. *BMC Syst. Biol.* **4,** 12 (2010).

65. Herrero, P., Fernández, R. & Moreno, F. Differential sensitivities to glucose and galactose repression of gluconeogenic and respiratory enzymes from Saccharomyces cerevisiae. *Arch. Microbiol.* **143,** 216–219

66. Soontorngun, N., Larochelle, M., Drouin, S., Robert, F. & Turcotte, B. Regulation of Gluconeogenesis in Saccharomyces cerevisiae Is Mediated by Activator and Repressor Functions of Rds2. *Mol. Cell. Biol.* **27,** 7895–7905 (2007).

67. Boubekeur, S., Bunoust, O., Camougrand, N., Castroviejo, M., Rigoulet, M. & Guérin, B. A Mitochondrial Pyruvate Dehydrogenase Bypass in the YeastSaccharomyces cerevisiae. *J. Biol. Chem.* **274,** 21044–21048 (1999).

68. Koukourakis, M. I., Giatromanolaki, A., Sivridis, E., Gatter, K. C. & Harris, A. L. Pyruvate Dehydrogenase and Pyruvate Dehydrogenase Kinase Expression in Non Small Cell Lung Cancer and Tumor-Associated Stroma. *Neoplasia* **7,** 1–6 (2005).

69. McFate, T., Mohyeldin, A., Lu, H., Thakar, J., Henriques, J., Halim, N. D., Wu, H., Schell, M. J., Tsang, T. M., Teahan, O., Zhou, S., Califano, J. A., Jeoung, N. H., Harris, R. A. & Verma, A. Pyruvate Dehydrogenase Complex Activity Controls Metabolic and Malignant Phenotype in Cancer Cells. *J. Biol. Chem.* **283,** 22700–22708 (2008).

70. Blows, M. W. & Hoffmann, A. A. A Reassessment of Genetic Limits to Evolutionary Change. *Ecology* **86,** 1371–1384 (2005).

71. Bonduriansky, R. & Day, T. Nongenetic Inheritance and Its Evolutionary Implications. *Annu. Rev. Ecol. Evol. Syst.* **40,** 103–125 (2009).

72. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).

73. Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Prados, J., Farinelli, L., Miska, E. & Mansuy, I. M. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat. Neurosci.* **17,** 667–669 (2014).

74. Zakrzewska, A., Eikenhorst, G. van, Burggraaff, J. E. C., Vis, D. J., Hoefsloot, H., Delneri, D., Oliver, S. G., Brul, S. & Smits, G. J. Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. *Mol. Biol. Cell* **22,** 4435–4446 (2011).

75. Durrant, W. E. & Dong, X. Systemic Acquired Resistance. *Annu. Rev. Phytopathol.* **42,** 185–209 (2004).

76. Dokudovskaya, S., Waharte, F., Schlessinger, A., Pieper, U., Devos, D. P., Cristea, I. M., Williams, R., Salamero, J., Chait, B. T., Sali, A., Field, M. C., Rout, M. P. & Dargemont, C. A Conserved Coatomer-related Complex Containing Sec13 and Seh1 Dynamically Associates With the Vacuole in Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **10,** M110.006478 (2011).

77. Lecointe, F., Simos, G., Sauer, A., Hurt, E. C., Motorin, Y. & Grosjean, H. Characterization of Yeast Protein Deg1 as Pseudouridine Synthase (Pus3) Catalyzing the Formation of $\Psi 38$ and $\Psi 39$ in tRNA Anticodon Loop. *J. Biol. Chem.* **273,** 1316–1323 (1998).

78. Amberg, D. C., Burke, D. & Strathern, J. N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. (CSHL Press, 2005).

79. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64,** 479–498 (2002).