GENOME ENGINEERING IN *SACCHAROMYCES CEREVISIAE*

BY

TONG SI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Huimin Zhao, Chair
Associate Professor Christopher V. Rao
Assistant Professor Brendan A. C. Harley
Assistant Professor Douglas A. Mitchell

Abstract

Microbial cell factory, which converts biomass feedstock to value-added compounds such as fuels, chemicals, materials and pharmaceuticals, has been proposed as a sustainable and renewable alternative to the traditional petrochemical industry. *Saccharomyces cerevisiae* is one of the most widely used microbial cell factories, to produce ethanol as the first generation of biofuel. To enable this yeast as a producer for 1-butanol, which is a next-generation gasoline substitute, I discovered, characterized and engineered an endogenous 1-butanol pathway in *S. cerevisiae*. Upon introduction of a single gene deletion *adh1*Δ, *S. cerevisiae* was able to accumulate more than 120 mg/L 1-butanol from glucose in the rich medium. Precursor feeding, ^{13}C-isotope labeling and gene deletion experiments demonstrated that the endogenous 1-butanol production was dependent on catabolism of threonine in a manner similar to fusel alcohol production by the Ehrlich pathway. Overexpression of the pathway enzymes and elimination of competing pathways achieved the highest reported 1-butanol titer in *S. cerevisiae* (243 mg/L).

Though 1-butanol titer was improved through pathway-based engineering, such rational design often meets with great challenges in cellular reprogramming due to our limited knowledge of complex biological systems. Directed evolution, on the other hand, has been proved as a better strategy by performing iterative cycles of mutagenesis and selection. Current practice of directed evolution is mostly confined to individual proteins, due to the lack of efficient tools to introduce mutations globally and iteratively in a genome. In the rest of this dissertation, I sought to develop a new method, RNA-interference assisted genome evolution

(RAGE), to apply directed evolution strategy in genome scale engineering of *S. cerevisiae*.

A functional RNA interference (RNAi) pathway was reconstituted in *S. cerevisiae* by introducing the Dicer and Argonaute proteins from *Saccharomyces castellii* as previously reported. We then performed the first RNAi screening in *S. cerevisiae*. The RNAi plasmid library was constructed with random genomic fragments and a convergent promoter expression cassette, which drives the *in vivo* synthesis of double-stranded RNAs (dsRNAs) to mediate knockdown of homologous genes. The library was confirmed with a complete coverage of the yeast genome, and employed to perform a suppressor analysis of a telomere-defect mutation *yku70*Δ. Two known and three novel knockdown modifications were identified to alleviate the growth arrest of the *Δyku70* strain at higher temperature, confirming the effectiveness of our RNAi library for genotyping. After establishing RNAi screening in *S. cerevisiae*, we combined it with directed evolution to rapidly engineer yeast cells for improved acetic acid (HAc) tolerance. Three rounds of iterative RNAi screening resulted in accumulation of three gene knockdown modifications that acted synergistically to confer substantially improved HAc tolerance. Together, these results demonstrated the RAGE method as an efficient, genome-scale and generally applicable strategy for directed genome evolution in *S. cerevisiae*.

I then expanded the application of RAGE to create a comprehensive genetic library (RAGE2.0). By directional cloning of a full-length, normalized cDNA library, one-step construction of the genome-wide ORF-overexpression and anti-sense RNA libraries was achieved. In the presence of the RNAi pathway, the RAGE2.0 library resulted in genome-wide

overexpression and knockdown modifications simultaneously. A wide range of phenotypes, including protein secretion, substrate utilization, and fuel molecule production, were screened with the RAGE2.0 library in a high-throughput manner. Both overexpression and knockdown targets were successfully identified to improve these phenotypes.

I further developed the RAGE3.0 method for automated genome engineering in yeast. Upon introduction of specific double-stranded breaks (DSBs) in the repetitive sequences by CRISPR nucleases, the genome-wide overexpression and knockdown cassettes in the RAGE2.0 library can be integrated into the genomic loci of repetitive sequences at high efficiency. This process can be iteratively performed to accumulate multiple genetic modifications in a single cell of an evolving yeast population. RAGE3.0 only involves simple liquid handling steps, hence it is readily automated with an integrated robotic platform. We envision this automated genome engineering method can enable generation of vast genetic diversity from which new or improved properties may emerge, and therefore greatly accelerate basic and applied biological research in *S. cerevisiae*.

Acknowledgements

I would like to thank my advisor, Professor Huimin Zhao, for his insight, advice and guidance, without which the accomplishment of these challenging and exciting projects will be impossible. I feel privileged to have such a caring and supportive advisor, who provides not only extraordinary freedom for me to pursue research directions that interest me most, but also his unconditional faith in me to succeed, which is important for a young student who constantly encounter frustration, doubts and setbacks. This feeling can be best expressed by a lyric quote—"you raise me up to more than I can be."

I am grateful to my dissertation committee members Professor Christopher V. Rao, Professor Brendan A. C. Harley and Professor Douglas A. Mitchell, for their insightful comments and suggestions, as well as support and advice on my academic career development.

I am also very indebted to every former and current group members of Zhao's group. Particularly, I would like to thank Dr. Fei Wen and Dr. Nikhil Nair for the initial training when I first joined the group. I also thank Dr. Jing Du, Dr. Yunzi Luo, Dr. Han Xiao, Ran Chao, Zehua Bao, Lu Zhang, Dr. Shuobo Shi, Dr. Youyun Liang and Dr. Jie Sun for fruitful collaboration on a few projects. I always benefits a lot for inspiring discussions with Dr. Jing Liang, Dr. Ning Sun, Dr. Meng Wang, Jiazhang Lian and Dr. Zengyi Shao. I also feel very lucky to meet with two visiting professor, Prof. Yu Jiang and Prof. Chun Li; we had a really good time. I want to thank the undergraduate students who has worked with me, including Sherrie Lim, Wei Yang and Alex Li. In addition, I really appreciate the help from Dr. Alexander Ulanov and Dr. Lucas Li at the Roy J. Carver Metabolomics Center for GC-MS and LC-MS analysis, as well as Dr. Barbara Pilas at the Roy J. Carver Biotechnology Center for flow cytometry experiments. Related to

academic career, I must thank Prof. Yanmei Li, my undergraduate advisor, who helped me to establish serious interest in research. I also want to acknowledge the support of University Block Grant, Mavis Future Faculty Fellowship and Drickamer Fellowship for financial support and related training opportunities.

The journey to a Ph.D. degree has never been easy, and I am really thankful for my dear friends to remind me there is a life beyond a lab bench. The Saturday nights in last five years will always be remembered, when I was "singing while enjoying the hot pot" with Weixin Tang, Yu Xiang, Yang Wang, Yi Yu, Yunzi Luo, Guodong Rao, Ke Ke, Yichen Tan, Bing Zuo, Cong Zhang, and Hang Xing. I appreciate all the wonderful moments with the Zhao group members outside the lab, such as when we were trapped on a snowy highway, when we were running full marathon, and when we were hiking backcountry trails. I would like to thank my fellow colleagues in the Chinese Student and Scholar Association and the Tsinghua University Alumni Association; together we have accomplished so much. I am also so fortunate to have my loyal friends, Xiang Li and Lijun Han; their unceasing support help me to walk through any difficult times.

I save my deepest thanks to my parents in the end, for their unconditional love, encourage and support to their only son who travels to the opposite side of this planet. Born in a family where both my parents worked at an ammonia plant, it is very interesting that I finally end up with a Ph.D. degree of Chemical Engineering. Therefore I would like to dedicate this dissertation to my dear parents who provide me with the Chemical Engineering related genes.

Table of Contents

Chapter 1 Introduction 1

1.1. Microbial cell factory 1

1.2. Genome scale engineering 2

1.2.1. Targeted genome editing 3

1.2.2. Transcriptome engineering 4

1.2.3. Genome synthesis 8

1.3. High-throughput genotyping with genetic libraries 8

1.4. RNA interference 11

1.5. Directed evolution 13

1.6. Project overview 15

1.7. References 18

Chapter 2 Endogenous 1-Butanol Production in *Saccharomyces cerevisiae* 27

2.1. Introduction 27

2.2. Results 30

2.2.1. ADH1 as a switch for 1-butanol production 30

2.2.2. Precursor feeding studies in the proposed 1-butanol pathway 34

2.2.3. Characterization of the 1-butanol pathway by gene deletion 41

2.2.4. Gene overexpression to improve 1-butanol production 42

2.2.5. Elimination of competing pathways 46

2.3. Conclusions and Future Prospects 47

2.4. Materials and methods 52

2.4.1. Media and cultivation conditions 52

2.4.2. DNA manipulation *53*

2.4.3. Strain constructions *54*

2.4.4. Metabolite detection *54*

2.5. References 58

Chapter 3 RAGE1.0: Complex Phenotype Engineering 63

3.1. Introduction 63

3.2. Results 67

3.2.1. Functional reconstitution of the RNAi pathway *67*

3.2.2. Suppressor analysis of the yku70Δ mutation by RNAi screening *72*

3.2.3. RNAi-assisted genome evolution for improved HAc tolerance *76*

3.2.4. Superior HAc tolerance of engineered strains by RAGE *81*

3.3. Conclusions and Future Prospects 83

3.4. Methods and Materials 89

3.4.1. Strains, media, and cultivation conditions *89*

3.4.2. DNA manipulation *90*

3.4.3. Reconstitution of the RNAi machinery in S. cerevisiae *91*

3.4.4. Construction of a genome-wide RNAi library *91*

3.4.5. Construction of a yeast knockdown library and screening *92*

3.4.6 Analysis of the newly discovered RNAi targets *93*

3.4.7. Second and third rounds of RAGE *94*

3.4.8. Characterization of engineered strains for HAc tolerance *95*

3.4.9. Estimation of the gene knockdown efficiency *96*

3.5. References 102

Chapter 4 RAGE 2.0: Constructing a Comprehensive Yeast Library.....109

4.1. Introduction 109

4.1.1. Comprehensive genome-wide library 109

4.1.2. Construction of a normalized, full-length cDNA library 110

4.2. Results 112

4.2.1. Directional cloning of a full-length, normalized cDNA library of S. cerevisiae 112

4.2.2. High-throughput screening with the RAGE2.0 library 115

4.3. Conclusions and Future Prospects..... 123

4.4. Materials and Methods 127

4.4.1. Strains, media, and cultivation conditions..... 127

4.4.2. DNA manipulation 127

4.4.3. Construction of the RAGE2.0 library 128

4.4.4. Screen of EGII-displaying mutants..... 129

4.4.5. Screen of glycerol-utilizing mutants..... 130

4.4.6. Screen of isobutanol-producing mutants 131

4.5. References 132

Chapter 5 RAGE3.0: Automated Genome Engineering in Yeast.....138

5.1. Introduction 138

5.2. Results 141

5.2.1. Design of RAGE3.0..... 141

5.2.2. Highly efficient δ-integration of a GFP reporter..... 144

5.2.3. RAGE3.0 to improve glycerol utilization 147

5.3. Conclusions and Future Prospects..... 150

5.4. Methods and Materials 153

5.4.1 Strains, media, and cultivation conditions *153*

5.4.2. DNA manipulation *154*

5.4.3. RAGE3.0 cycle: transformation, induction and selection *155*

5.4.4. Analysis of the GFP reporter system *156*

5.4.5. Screen of glycerol-utilizing mutants *157*

5.5. References 160

Chapter 1 Introduction

1.1. Microbial cell factory

Microbial cell factory, which converts biomass resources to value-added compounds such as fuels, chemicals, materials and pharmaceuticals, has been proposed as a sustainable and renewable alternative to the traditional petrochemical industry[1]. However, microorganisms are evolved to reproduce and survive in their native niche, and thus intensive reprogramming of the cellular machinery is needed to tailor the host cells for industrial applications[2, 3]. Numerous attempts have been made to overcome technical hurdles in the development of economically feasible processes based on microbial cell factories. For example, the cellulolytic capacity has been engineered into production hosts by secreting or surface-displaying cellulases/hemicellulases[4, 5]. Engineered metabolic pathways have been introduced to enable the fermentation of non-glucose sugars derived from lignocellulosic feedstock[6]. Novel biochemical transformations performed by heterologous pathways in either native or engineered forms can expand the scope of molecules that the microbes can synthesize. Both non-native pathways and host metabolism have been modified to redirect the metabolic fluxes without compromising the general fitness, in order to maximize the fermentation productivity and yield[6-10]. High levels of tolerance to the inhibitors in the lignocellulosic hydrolysate, the substrates and the final products as well as the harsh industrial environment (temperature, pH, osmotic pressure, *etc.*) are needed for robust production processes.

Though impressive progress has been made to engineer these industrially desired traits, the task of strain improvement remains challenging. This is because both our knowledge of complex biological systems and the tools to manipulate the cellular machinery are rather limited.

Conventional strain engineering approaches relies on random mutagenesis, which is achieved through chemical mutagens/UV irradiation[11], prolonged cultivation under selective pressure[12], transposon insertions[13-15] and genome shuffling[16, 17]. These methods are often labor-intensive, time-consuming, dependent on serendipity, and extremely difficult to analyze and transfer the genetic basis of a selected trait. Recently, the scale, efficiency and precision of genetic analysis and manipulation have been remarkably improved by several enabling technologies, including but not limited to microarray DNA synthesis, next-generation DNA sequencing (NGS), programmable DNA-binding proteins, in vivo biosensors, *etc*. Nowadays, billions of genome variants can be created in a directed and/or combinatorial manner, and the mutant strains with the optimal performance can be rapidly isolated. Collectively, these new technologies and their applications exemplify an emerging discipline called "genome scale engineering"[18-21].

1.2. Genome scale engineering

The practice of genome scale engineering can be broadly classified into three categories: genome editing, transcriptome engineering and genome synthesis (**Fig. 1.1**). Genome editing precisely or combinatorially modify the target genome at multiple loci. Modifications locate either in the open-reading frames (ORFs) or in the *cis-acting* regulatory elements such as promoters and ribosome-binding sites (RBSs). Transcriptome engineering essentially targets *trans-acting* regulatory elements, such as transcription factors (TFs) and non-coding RNAs (ncRNAs), by mutating endogenous regulators or introducing artificial ones. Genome synthesis involves hierarchical assembly of short chemically synthesized DNA fragments into genomic constructs. Although current synthetic genomes are constructed mainly based on their wild type templates, the ultimate goal is to write genome sequences *de novo*.

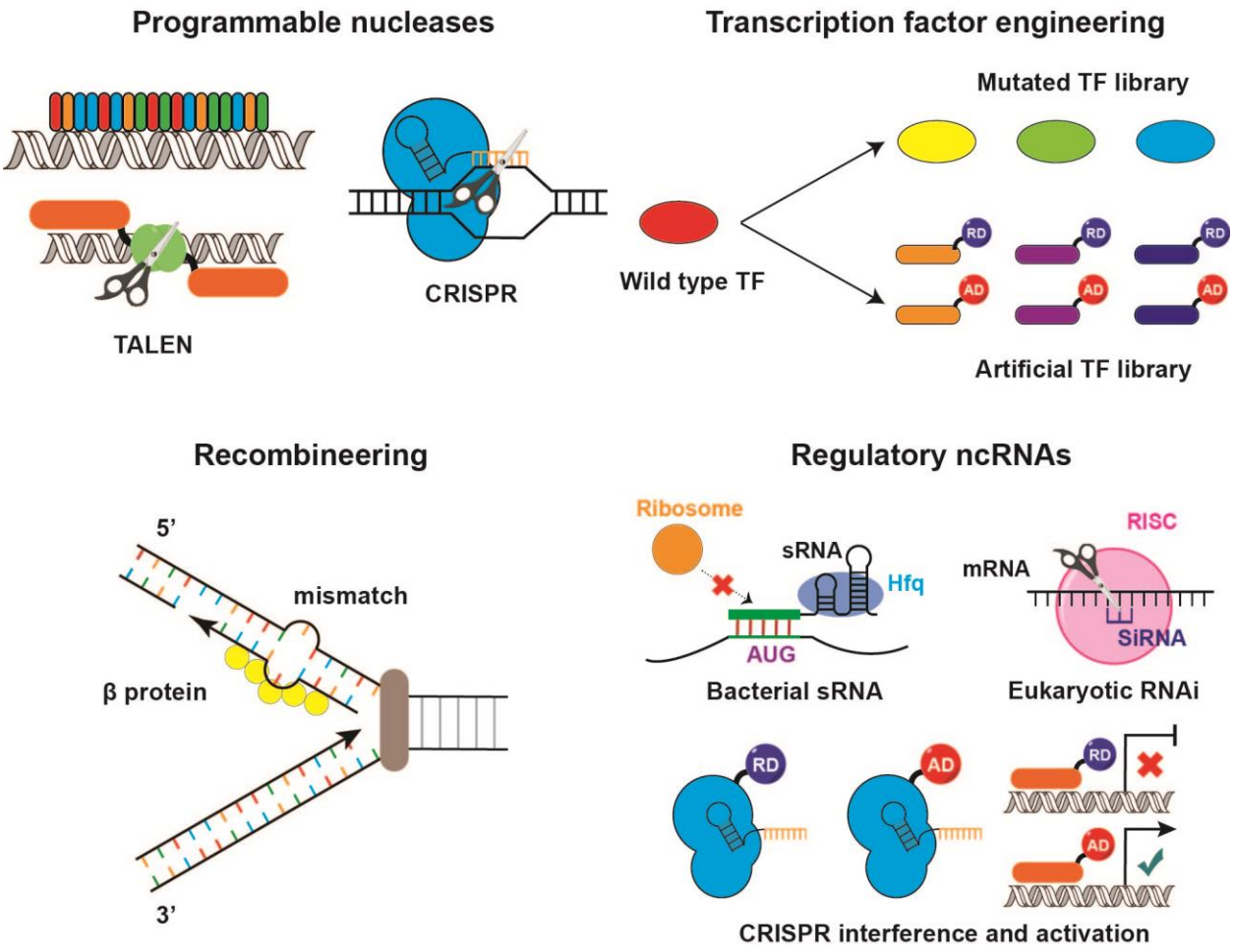

Figure 1.1. Overview of genome scale engineering tools in microbial systems. Programmable nucleases and recombineering are the two most promising targeted genome editing technologies. Transcriptome engineering targets *trans-acting* regulatory elements including transcription factors and non-coding RNAs.

1.2.1. Targeted genome editing

Targeted genome editing can be performed with a series of tools including recombinases[22, 23], group II intron retrotransposition[24], programmable nucleases[25], and recombineering[26, 27]. The last two tools are considered most promising due to their high targeting efficiency and ease of retargeting (**Fig. 1.1**).

Programmable nucleases mainly includes TALENs (Transcription Activator-Like Effector

Nucleases) and CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) nucleases, whose DNA-binding specificities can be easily predicted and altered[25]. These nucleases are introduced to create double-stranded breaks (DSBs) at target DNA sequences, and DSBs greatly stimulate homologous recombination (HR) to facilitate gene modification near the DSBs[20, 25, 28].

Recombineering uses phage proteins (RecET, λ-Red) to efficiently recombine a donor DNA strand with a bacterial genome via HR[26, 29]. In particular, recombination with ssDNA oligonucleotides can be mediated by the λ-Red-like protein in *Escherichia coli*, where oligos are incorporated to the lagging strand during genome replication at high efficiency[30]. Combining recombineering with microarray-derived oligonucleotide pools, the Multiplex Automated Genome Engineering (MAGE) method can continuously introduce combinatorial modifications across the *E. coli* genome[31].

1.2.2. Transcriptome engineering

Comparing with targeted genome editing, transcriptome engineering provides a complementary strategy for genome scale engineering. Targeting at *trans-acting* regulatory elements, genetic modulation is achieved without modifying the target chromosome loci. This feature not only eliminates the need for prior knowledge of host genomes, which is required by most genome editing methods that depend on HR, but also avoids the neighboring effects caused by physical modifications on a genome. Transcription factors (TFs) and regulatory non-coding RNAs (ncRNAs) are the most common targets for transcriptome engineering (**Fig. 1.1**).

Thanks to transcriptional regulatory networks, cells can rapidly coordinate the expression of thousands of genes when facing both internal and environmental stimuli[32]. Such networks

exhibit pyramid-shaped hierarchical structures, with most transcription factors (TF) at the bottom levels, some midlevel TFs, and only a few master TFs on top[33]. Whereas specific TFs at the bottom levels modulate dozens of genes in the same functional group, the master TFs have global influence over the gene expression profile[33]. Such properties make TFs the ideal targets for transcriptome reprogramming by modulating many genes simultaneously[34, 35].

Two main strategies have been applied to engineer TFs: modulation of native transcriptional machinery and introduction of artificial TFs. As a demonstration of native TF engineering, Global Transcriptional Machinery Engineering (gTME) introduces mutations to the master TFs that mainly mediate DNA recognition, based on the assumption that variations in these TFs may exert substantial changes to the promoter preference of the RNA polymerase. As proof of concept, the principal sigma factor in *E. coli* (σ^{70}) was subjected to error-prone PCR. From the resultant strain libraries, mutants with improved tolerance to sodium dodecyl sulfate (SDS) and ethanol were identified through serial subculturing[36]. In *Saccharomyces cerevisiae*, the TATA-binding protein Spt15p and a TATA-binding protein-associated factor Taf25p were mutated. The best variant, which harbored three amino acid mutations in Spt15p, conferred a 70% improvement in the ethanol productivity[37]. On the other hand, artificial transcription factor (ATF) libraries have also been created to generate transcriptional diversity. A minimal ATF may only contain a DNA-binding domain, whose interaction with its target sequence most likely down-regulate the expression of a nearby gene by interfering with the transcriptional initiation or elongation[38]. The DNA-binding domain can also be attached to effector (activator/repressor) domains or ligand-binding domains, which permits more sophisticated regulation. Most ATF reported so far has employed zinc finger proteins (ZFPs) as the DNA-binding domains, and the library of ATFs is constructed through combinatorial assembly of individual zing fingers with

diverse DNA-binding specificities. The first example of such effort was the introduction of over 10^5 ZFPs attached with effector domains into *S. cerevisiae*[39]. The library consisted of three- or four-finger ZFPs, which recognize 9bp or 12bp DNA sequences with limited randomness constrained by choice of individual fingers (40 and 25 individual fingers for three- and four-finger proteins, respectively). Several artificial TFs have been identified to confer a number of tolerance phenotypes towards heat, osmotic pressure and an antifungal drug ketoconazole[39]. A similar strategy has been applied to *E. coli* to isolate tolerant strains towards heat shock[38] and butanol[40].

In addition to TF networks, non-coding RNA molecules (ncRNAs) are also increasingly recognized as key regulators across the biological kingdoms[41, 42]. To be suitable for genome-wide application, synthetic ncRNAs should be preferably *trans-acting*, permitting simple introduction of a genome-wide library with minimal considerations on local genetic context. Also, the interaction between an ncRNA and its DNA or mRNA target should be mainly determined by Watson-Crick base pairing, so that the binding specificity and efficiency can be predictable and programmable.

In bacteria, *trans-acting* small RNAs (sRNAs) and antisense RNAs (asRNAs) are two main regulatory ncRNAs[41]. Na et al. recently developed a general framework to design synthetic sRNAs in *E. coli* for metabolic engineering[43]. The synthetic sRNAs were composed of a scaffold sequence and a target-binding sequence. The scaffold was derived from a naturally occurring sRNAs, *MicC*, and the scaffold can recruit the Hfq protein to facilitate sRNA-mRNA interaction and mRNA degradation. The native target binding sequence of *MicC* can be replaced by the antisense sequence to the translation initiation region (TIR) of any given gene. Correlation was found between the repression capability and the binding energy of the antisense

sequence, which allowed for fine-tuning of knockdown efficiency. Although the sRNA library constructed in that study only target the cadaverine production related genes[43], it is possible to expand the strategy to a genome scale.

On the other hand, asRNAs have been used for functional genomics study in a series of bacteria, such as *Streptococcus mutans*[44] and *Staphylococcus aureus*[45]. However, it has been long recognized that asRNAs are inefficient for gene repression in *E. coli*[46]. Recently, it was found that asRNA molecules with paired-termini have enhanced stability and hence improved repression capacity[47]. Enzymatic digested *E. coli* genomic DNA fragments has been cloned into a paired-termini expression vector to generate a genome-wide asRNA library, which was used to successfully isolate asRNAs that target essential genes and led to conditional growth inhibitory[48].

As for eukaryotes, the most common ncRNA machinery for gene expression regulation is RNA interference (RNAi), a cellular gene silencing mechanism whereby mRNAs are targeted for degradation by homologous double-stranded RNAs (dsRNAs)[49, 50]. RNAi proves to be a powerful tool for genome-wide reduction-of-function screen in many higher eukaryotes[51, 52], yet its applications in microbes are rare. This is probably due to the lack of a native RNAi pathway in *S. cerevisiae*, which is the most-widely used microbial eukaryote. Recently, heterologous RNAi machinery has been reconstituted in *S. cerevisiae*[53], which opens up the possibility of genome-wide RNAi screen in this yeast.

In addition to naturally-occurring ncRNAs, synthetic RNAs were also used to modulate gene expression in the CRISPR-mediated interference (CRISPRi) and CRISPR-mediated activation technologies. By co-expressing of a Cas9 mutant with abolished endonuclease activity (dCas9) and a guide RNA targeting at the non-template DNA strand of a target gene, up to 1,000-fold reduction in gene expression was achieved in *E. coli*[54]. In *S. cerevisiae*, the silencing efficiency

may be further improved by fusing the dCas9 protein to transcription repressors or chromatin silencers[55]. For gene activation, transcriptional activators can be delivered by the dCas9-guide RNA complex to the upstream region of a promoter, resulting in up-regulation in *E. coli*[56] and yeast[55, 57].

1.2.3. Genome synthesis

Genome synthesis is one of the most impressive achievement of synthetic biology, ranging from viral genomes[58, 59], bacterial genomes[60-62] to yeast chromosomes[63, 64]. Early efforts mainly focused on increasing the scale of the final DNA constructs, from the 7.5 kb cDNA copy of a poliovirus genome[58] to a 1.08 Mb bacterial genome[61]. In terms of DNA sequences, the synthetic genomes are almost exact copies of the native ones, except a few inserted "watermarks" such as the names of the team members[60, 61]. A recent report took a step further to build a designer yeast chromosome that was substantially different from its wild-type template[64]. Compared to the native chromosome III of *S. cerevisiae*, the designer chromosome synIII was ~ 14% smaller due to the deletion of some non-essential regions such as transfer RNAs, transposons, introns, *etc*. However, the design principles are still very simple, and it requires substantial technological development before we can write a fully synthetic genome.

1.3. High-throughput genotyping with genetic libraries

One important application of genome scale engineering is to create genome-wide libraries, which are powerful tools to comprehensively investigate the impact of individual genetic modification on a given phenotype. For overexpression libraries, either genomic fragments[65] or

all the open reading frames (ORFs) under the control of a promoter[66] can be cloned into an extrachromosomal vector. After screen/enrichment, the inserts can be identified through microarray analysis[67] or DNA sequencing. Genome-wide knockout libraries can be generated by many ways. Transposon mutagenesis has been optimized for unbiased integration of an antibiotic marker cassette into the entire genome, hence creating a random knockout library[68, 69]. Also, all the nonessential genes can be disrupted through homologous recombination, resulting in the construction of the yeast deletion collection[70, 71] and the Keio knockout collection in *E. coli*[72]. In addition to knockout, reduction-of-function screen has also been applied for genome-wide analysis. Examples include screening with asRNAs and RNAi that are discussed before. Knockdown libraries are especially important to study essential genes whose deletion mutations are lethal. For example, insertion of an antibiotic marker into the terminator region to destabilize the mRNA enabled knockdown modification on the essential genes in *S. cerevisiae*[73]. Notably, there are also technologies that can create comprehensive genetic libraries including both overexpression and knockdown modifications. For the trackable multiplex recombineering (TRMR), two kinds of synthetic cassettes were designed for promoter replacement: the 'up' cassette containing a strong promoter, and the 'down' cassette containing an inert sequence to replace the native RBS. Through recombineering, these synthetic cassettes were incorporated in front of every gene in *E. coli*, which led to either increased or decreased expression of a target gene[74].

To facilitate subsequent analysis with these genetic libraries, molecular barcodes have been used to monitor the abundance of every mutant strain in a mixed population. Microarray analysis with complementary probes[75], as well as the "Bar-Seq" method[76], can be used to quantify the dynamics of barcodes and their linked mutants in various screen experiments,

enabling high-throughput mapping of relevant genes to a given phenotype. Such high-throughput capacity explains the wide application of molecular barcodes in analyzing overexpression[66], knockout[70] and TRMR libraries[74].

Adding more dimensions to such approaches, i.e. modification at two or even more loci, is necessary because of the non-linear interactions between single genetic variations. For combinatorial overexpression libraries, coexpressing genomic libraries (CoGEL) constructed genomic libraries in a series of vectors (plasmid or fosmid) with compatible replication origins and different resistance markers, which enables coexistence of two or more genomic insets in one cell. This approach successfully identified known and novel combinations of genetic changes that conferred improved acid tolerance in *E. coli*[77]. On the other hand, construction of double-mutant library from single loss-of-function collections by mating or conjugation has been demonstrated in model organisms such as *E. coli*[78, 79] and *S. cerevisiae*[80, 81]. An impressive application was the depiction of a genome-scale digenic interaction network in *S. cerevisiae*, by examining 5.4 million gene-gene pairs in double-mutant library[82]. However, current protocols to generate genome-wide double-mutant libraries are quite resource-intensive and time-consuming, as complicated replica-pinning procedures are needed to perform mating, recombination and selection. Therefore, alternative approaches that simplify the introduction of a second mutation on a genome scale are desirable to speed up the discovery of synergistic modifications. The introduction of an RNAi library to create genome-wide knockdown modifications requires only a single step of transformation of the host cells[83, 84]. Such simplicity and effectiveness should enable the use of iterative RNAi screening to accumulate beneficial knockdown modifications.

1.4. RNA interference

Since its discovery in *Caenorhabditis elegans* in 1998[49], RNA interference (RNAi) has literally "heralded a revolution" in biotechnology[50]. RNAi is a cellular gene silencing mechanism, where messenger RNAs are targeted for degradation by homologous double-stranded RNA (dsRNA). The major functions of RNAi include post-transcriptional modulation of gene expression, assembly of silent chromatin structures, and defense against invasion by heterologous nucleic acids such as viruses[50].

Different organisms comprise different mechanisms for RNAi, but the basic process shared three common steps (**Fig. 1.2**)[50]. First, small RNA duplexes (siRNAs) are generated from long dsRNA precursors by a ribonuclease III (RNase III) enzyme called Dicer. Second, these small RNAs are loaded into Argonaute proteins to form a protein-RNA complex known as RNA-induced silencing complex (RISC). Third, RISC finds and cleaves the target mRNA molecule whose sequence is homologous to the small RNA loaded in the complex. In this way, sequence-specific gene silencing can be achieved by dsRNAs and the RNAi pathway. The effectiveness and specificity of RNAi facilitate its wide use in functional genomics, therapeutics and metabolic engineering. RNAi-based genetic screens have been demonstrated in both lower organisms and mammalian cells[51, 52], which help to elucidate various important biological mechanisms underlying developmental biology, signal transduction pathways and human diseases.

Successful RNAi screening experiments depend on a well-designed library of RNAi reagents and high-throughput read-out of target phenotypes[51, 52]. A collection of RNAi reagents, synthetic or derived from genomic DNA/cDNA, has been constructed to suit different purposes of genetic screens. RNAi reagents can be generated *in vitro* and then delivered into the cells transiently, or can be generated *in vivo* from vectors for long-term effects. On the other hand, two screening

formats can be applied for phenotype characterization. Either a systematic screen where each gene is silenced individually, or a pooled library coupled with high-throughput selection and screening can be used to target many genes at once. While the systematic screen permits a comprehensive depiction of the genotype-phenotype relationship for a given trait, the selection strategy provides a higher throughput and less expensive method based on the selectable cell growth advantage or fluorescence-activated cell sorting (FACS). Since its first demonstration in *C. elegans* in 2000, RNAi screening has accumulated a multitude of tools to expand the power of this genome engineering strategy.

Recently, a functional RNAi pathway has been reconstituted in *S. cerevisiae* by introducing two RNAi proteins Ago1 and Dcr1 from another budding yeast *Saccharomyces castellii*[53]. However, RNAi screening has yet been reported in *S. cerevisiae*. Though there are several genome-wide libraries available for *S. cerevisiae*, RNAi screening possesses some unique advantages over the classical loss-of-function screens. First, for those essential genes whose deletion is lethal and thus not accessible in gene knockout libraries, RNAi screening provides a way to assay their functions by creating reduction-of-function mutants that are still viable. Second, RNAi screening enables the fine-tuning of gene expression by knockdown, which theoretically provides more information than comparing only two states: wild type and complete deletion. Third, a modifier screening can provide invaluable insight about genetic interactions by performing a genome-wide screening in a mutant background, but it requires a time-consuming and labor-intensive process to cross a query strain with a complete gene deletion library. On the contrary, a pooled RNAi library can be introduced through a single step of transformation in a query background, which will greatly save the time and labor to carry out the modifier screening. Based on these considerations, it is highly desirable to devise an RNAi screening platform in *S.*

cerevisiae.

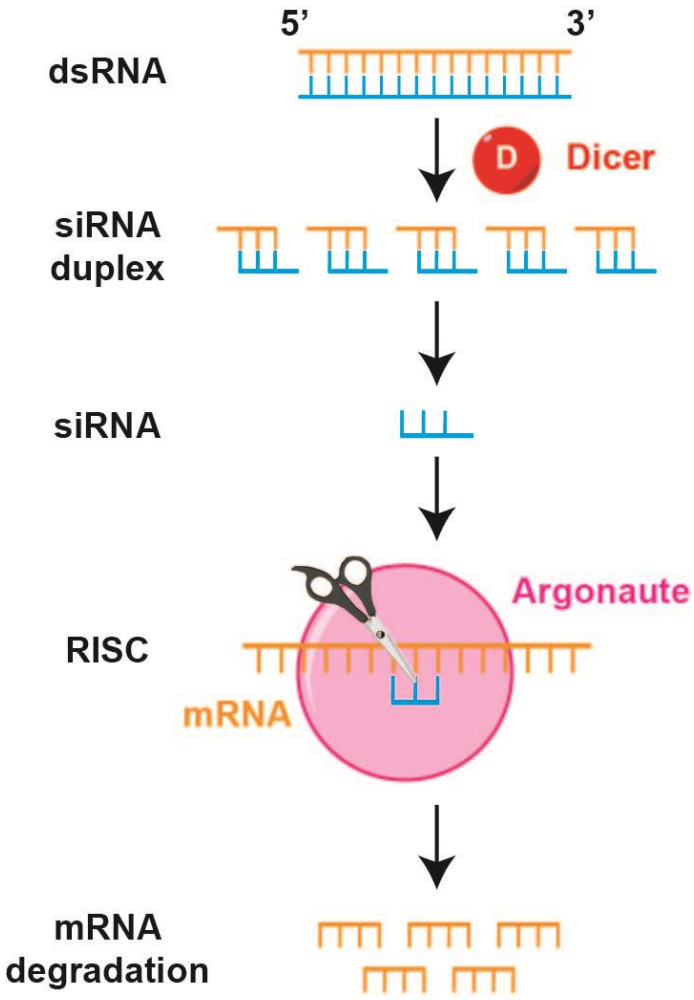

Figure 1.2. Scheme of RNA interference process. Long ds long dsRNA precursors are digested by the Dicer protein to produce small interference RNA duplexes (siRNAs). Then, siRNAs are loaded into Argonaute proteins to form RNA-induced silencing complex (RISC). RISC finds and cleaves the target homologous mRNA molecule, and leads to gene silencing.

1.5. Directed evolution

There are two methods to engineer biological entities: rational design and directed evolution. Given our limited knowledge of the complex biological systems, directed evolution has been

proved as a better engineering strategy than rational design[18]. By mimicking the natural evolution process—variation and selection—in the laboratory[85], directed evolution has achieved enormous success in engineering industrially relevant properties[5, 86, 87], without the need for the prior knowledge of target systems. However, the success of directed evolution is mostly confined to the level of individual proteins. The main obstacle to apply this strategy on a genome scale is the lack of efficient tools to create genome-wide diversity both globally and iteratively.

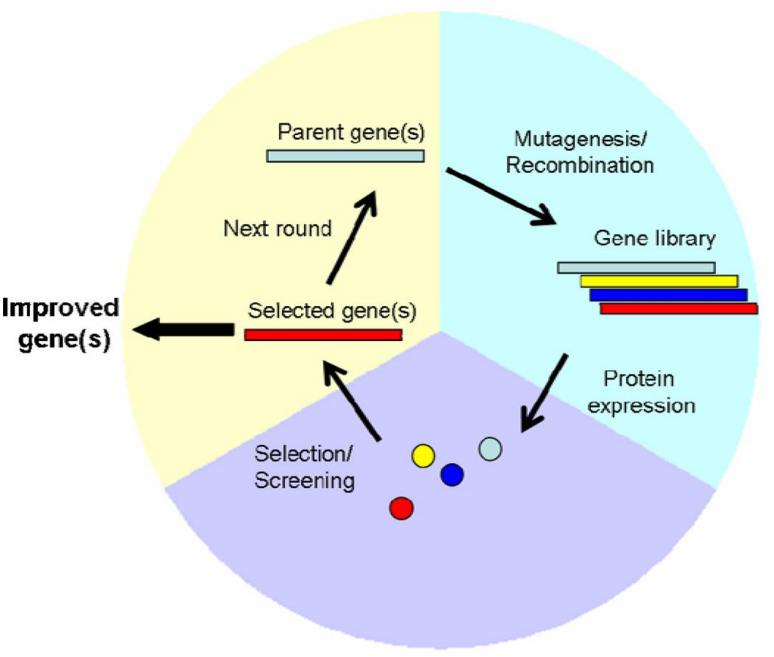

Figure 1.3. Scheme of directed protein evolution[5]. Iterative cycles of mutagenesis and selection/screening are performed to continuously improve the properties of the target protein, until the engineering objectives are achieved or no more improvement is observed.

To design a feasible strategy for directed genome evolution, it is necessary to learn from the success of directed protein evolution. To identify variants with altered or improved functions, four key steps are performed (**Fig. 1.3**)[87]: First, choosing a good starting parent; second, creating a library of variants; third, selecting variants with desired property and last, repeating the process

until no further improvement can be achieved or the engineering objective is reached. While different selection/screen strategies are needed for a variety of applications, the creation of library generally falls into two categories[85]. The first method is "asexual" evolution by sequential rounds of random mutagenesis, and the other method is "sexual" evolution by gene recombination. Error-prone PCR is the standard method to introduce point mutations randomly into a protein sequence, and DNA shuffling is the most common strategy to create combinations of selected mutations. These mature techniques provide convenient ways to sample the protein sequence space by continuously walking through the "fitness landscapes." Library construction techniques that can fulfill similar tasks on a genome scale are definitely needed to enable directed genome evolution.

1.6. Project overview

This dissertation focuses on the engineering of *S. cerevisiae* as an efficient microbial cell factory for chemical and fuel production. Two different approaches were taken: the traditional metabolic engineering strategy to improve the endogenous 1-butanol production; and a new method I developed named RNAi-Assisted Genome Evolution (RAGE) for genome-scale engineering.

Chapter 2 reports the discovery, characterization and engineering of an endogenous 1-butanol pathway in *Saccharomyces cerevisiae*. Upon introduction of a single gene deletion *adh1*Δ, *S. cerevisiae* was able to accumulate more than 120 mg/L 1-butanol from glucose in rich medium. Precursor feeding, ^{13}C-isotope labeling and gene deletion experiments demonstrated that the endogenous 1-butanol production was dependent on catabolism of threonine in a manner similar

to fusel alcohol production by the Ehrlich pathway. Specifically, the leucine biosynthesis pathway was engaged in the conversion of key 2-keto acid intermediates. Metabolic engineering efforts, including overexpression of the pathway enzymes and elimination of competing pathways, achieved the highest reported 1-butanol titer in *S. cerevisiae* (242.8 mg/L).

In Chapter 3, I established RAGE as a generally applicable method for genome scale engineering in the yeast *S. cerevisiae*. Through iterative cycles of creating a library of RNAi induced reduction-of-function mutants coupled with high throughput screening or selection, RAGE can continuously improve target trait(s) by accumulating multiplex beneficial genetic modifications in an evolving yeast genome. To validate the RNAi library constructed with yeast genomic DNA and convergent-promoter expression cassette, I demonstrated RNAi screening in *S. cerevisiae* for the first time by identifying two known and three novel suppressors of a telomerase-deficient mutation yku70Δ. I then showed the application of RAGE for improved acetic acid tolerance, a key trait for microbial production of chemicals and fuels. Three rounds of iterative RNAi screening led to the identification of three gene knockdown targets that acted synergistically to confer an engineered yeast strain with substantially improved acetic acid tolerance.

Chapter 4 describes the development of a comprehensive genetic library in yeast (RAGE2.0). I created a full-length cDNA library of *S. cerevisiae*, whose quality and coverage are improved through size fractionation and normalization. Directional cloning of the resultant cDNA collection achieved one-step construction of the ORF-overexpression and anti-sense RNA libraries. In the presence of the RNAi pathway, the RAGE2.0 library resulted in genome-wide overexpression and knockdown modifications simultaneously. A wide range of phenotypes, including protein secretion, substrate utilization, and fuel molecule production, were screened

with the RAGE2.0 library in a high-throughput manner. Both overexpression and knockdown targets were successfully identified to improve these phenotypes.

In Chapter 5, the RAGE3.0 method was devised for automated yeast genome engineering. An inducible CRISPR system was constructed to introduce specific double-stranded breaks (DSBs) in the repetitive δ sequences of the yeast genomes. The efficiency of integration of a GFP donor cassette into the δ elements was greatly improved; about 70% cells can be modified in a single round of transformation. CRISPR-assisted δ-integration was used to introduce the genome-wide overexpression and knockdown cassettes in the RAGE2.0 library, and this process was iteratively performed to accumulate dozens of genetic modifications in a single cell of an evolving yeast population. Mutant strains with multiplex genetic modifications were isolated with improved glycerol utilization. RAGE3.0 only involves simple liquid handling steps, hence it is readily automated with an integrated robotic platform in the future.

1.7. References

1. Zhao, H. & Chen, W. Chemical biotechnology: microbial solutions to global change. Editorial overview. *Curr Opin Biotechnol* **19**, 541-543 (2008).
2. Lee, J.W. et al. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol* **8**, 536-546 (2012).
3. Keasling, J.D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358 (2010).
4. Sun, J., Wen, F., Si, T., Xu, J.H. & Zhao, H. Direct conversion of xylan to ethanol by recombinant *Saccharomyces cerevisiae* strains displaying an engineered minihemicellulosome. *Appl Environ Microbiol* **78**, 3837-3845 (2012).
5. Wang, M., Si, T. & Zhao, H. Biocatalyst development by directed evolution. *Bioresour Technol* **115**, 117-125 (2012).
6. Du, J., Yuan, Y., Si, T., Lian, J. & Zhao, H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res* **40**, e142 (2012).
7. Ajikumar, P.K. et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74 (2010).
8. Tyo, K.E., Ajikumar, P.K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat Biotechnol* **27**, 760-765 (2009).
9. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).
10. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of

intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**, 1027-1032 (2006).

11. Crook, N. & Alper, H.S. in Engineering complex phenotypes in industrial strains 1-33 (John Wiley & Sons, Inc., 2012).
12. Portnoy, V.A., Bezdan, D. & Zengler, K. Adaptive laboratory evolution - harnessing the power of biology for metabolic engineering. *Curr Opin Biotechnol* **22**, 590-594 (2011).
13. Hamer, L., DeZwaan, T.M., Montenegro-Chamorro, M.V., Frank, S.A. & Hamer, J.E. Recent advances in large-scale transposon mutagenesis. *Curr Opin Chem Biol* **5**, 67-73 (2001).
14. Eckert, S.E. et al. Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of *Escherichia coli* O157:H7 screened in cattle. *J Bacteriol* **193**, 1771-1776 (2011).
15. Hutchison, C.A. et al. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165-2169 (1999).
16. Zhang, Y.X. et al. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415**, 644-646 (2002).
17. Biot-Pelletier, D. & Martin, V.J.J. Evolutionary engineering by genome shuffling. *Appl Microbiol Biotechnol* **98**, 3877-3887 (2014).
18. Carr, P.A. & Church, G.M. Genome engineering. *Nat Biotechnol* **27**, 1151-1162 (2009).
19. Esvelt, K.M. & Wang, H.H. Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol* **9**, 17 (2013).
20. Segal, D.J. & Meckler, J.F. Genome engineering at the dawn of the golden age. *Annu Rev Genomics Hum Genet* **14**, 135-158 (2013).

21. Jeong, J., Cho, N., Jung, D. & Bang, D. Genome-scale genetic engineering in *Escherichia coli*. *Biotechnol Adv* **31**, 804-810 (2013).

22. Turan, S., Zehe, C., Kuehle, J., Qiao, J. & Bode, J. Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene* **515**, 1-27 (2013).

23. Karow, M. & Calos, M.P. The therapeutic potential of PhiC31 integrase as a gene therapy system. *Expert Opin Biol Ther* **11**, 1287-1296 (2011).

24. Michel, F. & Ferat, J.L. Structure and activities of group II introns. *Annu Rev Biochem* **64**, 435-461 (1995).

25. Kim, H. & Kim, J.S. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* **15**, 321-334 (2014).

26. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* **20**, 123-128 (1998).

27. Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. & Court, D.L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4**, 206-223 (2009).

28. Sander, J.D. & Joung, J.K. CRIRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355 (2014).

29. Murphy, K.C. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**, 2063-2071 (1998).

30. Yu, D., Sawitzke, J.A., Ellis, H. & Court, D.L. Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc Natl Acad Sci U S A* **100**, 7207-7212 (2003).

31. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-U133 (2009).

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

33. Yu, H. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc Natl Acad Sci U S A* **103**, 14724-14731 (2006).

34. Lin, Z., Zhang, Y. & Wang, J. Engineering of transcriptional regulators enhances microbial stress tolerance. *Biotechnol Adv* **31**, 986-991 (2013).

35. Santos, C.N. & Stephanopoulos, G. Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr Opin Chem Biol* **12**, 168-176 (2008).

36. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* **9**, 258-267 (2007).

37. Alper, H., Moxley, J., Nevoigt, E., Fink, G.R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565-1568 (2006).

38. Park, K.S., Jang, Y.S., Lee, H. & Kim, J.S. Phenotypic alteration and target gene identification using combinatorial libraries of zinc finger proteins in prokaryotic cells. *J Bacteriol* **187**, 5496-5499 (2005).

39. Park, K.-S. et al. Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol* **21**, 1208-1214 (2003).

40. Lee, J.Y., Yang, K.S., Jang, S.A., Sung, B.H. & Kim, S.C. Engineering butanol-tolerance in escherichia coli with artificial transcription factor libraries. *Biotechnol Bioeng* **108**,

742-749 (2011).

41. Qi, L.S. & Arkin, A.P. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat Rev Microbiol* **12**, 341-354 (2014).

42. Kang, Z. et al. Small RNA regulators in bacteria: powerful tools for metabolic engineering and synthetic biology. *Appl Microbiol Biotechnol* **98**, 3413-3424 (2014).

43. Na, D. et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat Biotechnol* **31**, 170-174 (2013).

44. Wang, B. & Kuramitsu, H.K. Inducible antisense RNA expression in the characterization of gene functions in *Streptococcus* mutans. *Infect Immun* **73**, 3568-3576 (2005).

45. Forsyth, R.A. et al. A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol Microbiol* **43**, 1387-1400 (2002).

46. Wagner, E.G. & Flardh, K. Antisense RNAs everywhere? *Trends Genet* **18**, 223-226 (2002).

47. Nakashima, N., Tamura, T. & Good, L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in *Escherichia coli*. *Nucleic Acids Res* **34**, e138 (2006).

48. Meng, J. et al. A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing *Escherichia coli* essential genes. *FEMS Microbiol Lett* **329**, 45-53 (2012).

49. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811 (1998).

50. Hannon, G.J. RNA interference. *Nature* **418**, 244-251 (2002).

51. Boutros, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nat*

Rev Genet **9**, 554-566 (2008).

52. Echeverri, C.J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7**, 373-384 (2006).

53. Drinnenberg, I.A. et al. RNAi in budding yeast. *Science* **326**, 544-550 (2009).

54. Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).

55. Gilbert, L.A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451 (2013).

56. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429-7437 (2013).

57. Farzadfard, F., Perli, S.D. & Lu, T.K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth Biol* **2**, 604-613 (2013).

58. Cello, J., Paul, A.V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016-1018 (2002).

59. Chan, L.Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol Syst Biol* **1**, 2005 0018 (2005).

60. Gibson, D.G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215-1220 (2008).

61. Gibson, D.G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56 (2010).

62. Karas, B.J., Tagwerker, C., Yonemoto, I.T., Hutchison, C.A., 3rd & Smith, H.O. Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. *ACS Synth Biol* **1**, 22-28 (2012).

63. Dymond, J.S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).

64. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58 (2014).

65. Lynch, M.D., Warnecke, T. & Gill, R.T. SCALEs: multiscale analysis of library enrichment. *Nat Methods* **4**, 87-93 (2007).

66. Ho, C.H. et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol* **27**, 369-377 (2009).

67. Lynch, M.D., Gill, R.T. & Stephanopoulos, G. Mapping phenotypic landscapes using DNA micro-arrays. *Metab Eng* **6**, 177-185 (2004).

68. Badarinarayana, V. et al. Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat Biotechnol* **19**, 1060-1065 (2001).

69. Alexeyev, M.F. & Shokolenko, I.N. Mini-Tn10 transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of gram-negative bacteria. *Gene* **160**, 59-62 (1995).

70. Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391 (2002).

71. Winzeler, E.A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).

72. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).

73. Breslow, D.K. et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* **5**, 711-718 (2008).

74. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**, 856-862 (2010).

75. Pierce, S.E. et al. A unique and universal molecular barcode array. *Nat Methods* **3**, 601-603 (2006).

76. Smith, A.M. et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res* **19**, 1836-1842 (2009).

77. Nicolaou, S.A., Gaida, S.M. & Papoutsakis, E.T. Coexisting/Coexpressing Genomic Libraries (CoGeL) identify interactions among distantly located genetic loci for developing complex microbial phenotypes. *Nucleic Acids Res* **39**, e152 (2011).

78. Typas, A. et al. High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat Methods* **5**, 781-787 (2008).

79. Butland, G. et al. eSGA: *E. coli* synthetic genetic array analysis. *Nat Methods* **5**, 789-795 (2008).

80. Tong, A.H. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364-2368 (2001).

81. Pan, X. et al. A robust toolkit for functional profiling of the yeast genome. *Mol Cell* **16**, 487-496 (2004).

82. Costanzo, M. et al. The genetic landscape of a cell. *Science* **327**, 425-431 (2010).

83. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-437 (2004).

84. Echeverri, C.J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7**, 373-384 (2006).

85. Arnold, F.H. Design by directed evolution. *Accounts Chem Res* **31**, 125-131 (1998).

86. Cobb, R.E., Si, T. & Zhao, H. Directed evolution: an evolving and enabling synthetic biology tool. *Curr Opin Chem Biol* **16**, 285-291 (2012).

87. Romero, P.A. & Arnold, F.H. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* **10**, 866-876 (2009).

Chapter 2 Endogenous 1-Butanol Production in *Saccharomyces cerevisiae*

2.1. Introduction

Microbial cell factory, which offers sustainable solutions to global warming and energy crisis, has emerged as a promising alternative to the traditional petrochemical industry[1, 2]. Among the most prominent examples of microbial cell factory is the yeast *Saccharomyces cerevisiae*, which is currently used to produce ethanol as an alternative fuel[3]. However, due to their superior fuel characteristics compared to ethanol, advanced biofuels have attracted a growing interest[4, 5]. In particular, 1-butanol is considered a substantially better gasoline substitute than ethanol[6]. The energy density of 1-butanol (29.2 MJ/L) is comparable to that of gasoline (32.5 MJ/L), and much higher than that of ethanol (21.2 MJ/L)[7]. While ethanol can only be blended with gasoline to a final volume percentage of 85%, 1-butanol can be used in pure form or blended in gasoline at any ratio[7]. Furthermore, 1-butanol is more hydrophobic and less corrosive than ethanol, and therefore can be transported through the existing pipeline infrastructure[7].

Traditionally, biological synthesis of 1-butanol is performed by *Clostridium* species through the acetone-butanol-ethanol (ABE) fermentation process[8]. However, inherent challenges of *Clostridium* manipulation and fermentation, such as a lack of genetic tools, unavoidable byproducts and intolerance to 1-butanol and oxygen, have hindered further improvement of the ABE fermentation. Therefore, production of 1-butanol in industrially friendly organisms, such as *Escherichia coli* and *S. cerevisiae*, has been the focus of recent efforts[5]. Two metabolic routes

have been explored to produce 1-butanol. The first route is the heterologous expression of the *Clostridium* 1-butanol pathway[9, 10], which is essentially the same as the reversed β-oxidation pathway for butanol production[11]. This CoA-dependent pathway enables an impressive production titer (30 g/L) in *E. coli*[12], but its performance in *S. cerevisiae* is far less promising (2.5 mg/L[13] and 16.3 mg/L[14]). The second route takes advantage of the amino-acid metabolic pathways [15], where 1-butanol is produced via keto-acid intermediates in *E. coli*. The drawback of this strategy is the unavoidable co-production of 1-propanol, the synthesis of which shares a common intermediate (2-ketobutyrate) with the 1-butanol pathway[16].

Although the current production level of 1-butanol in *S. cerevisiae* is far behind *Clostridium* and *E. coli*[14], there are advantages to utilize *S. cerevisiae* as a 1-butanol producer. *S. cerevisiae* is very robust toward inhibitors and low pH condition, and there is no phage contamination issues with yeast compared to bacterial hosts[3]. *S. cerevisiae* has also been shown as one of the most tolerant organisms toward 1-butanol[17, 18]. Additionally, with the widest application in biorefinery, *S. cerevisiae* can serve as a "drop-in" cell factory with compatibility to current industrial infrastructure[19]. Low concentrations of 1-butanol have been found in fusel oil isolated from yeast wine fermentation[20], and the origin of 1-butanol is believed to be from plant-derived 2-keto acid precursors[21] or bacterial contamination[22]. Early studies showed 1-butanol accumulation in mutated isoleucine- and valine-requiring *S. cerevisiae* strains[22], implying a native 1-butanol pathway may exist in *S. cerevisiae* strains with certain genetic background. However, it is not clear whether endogenous 1-butanol production is a universal capacity of *S. cerevisiae*.

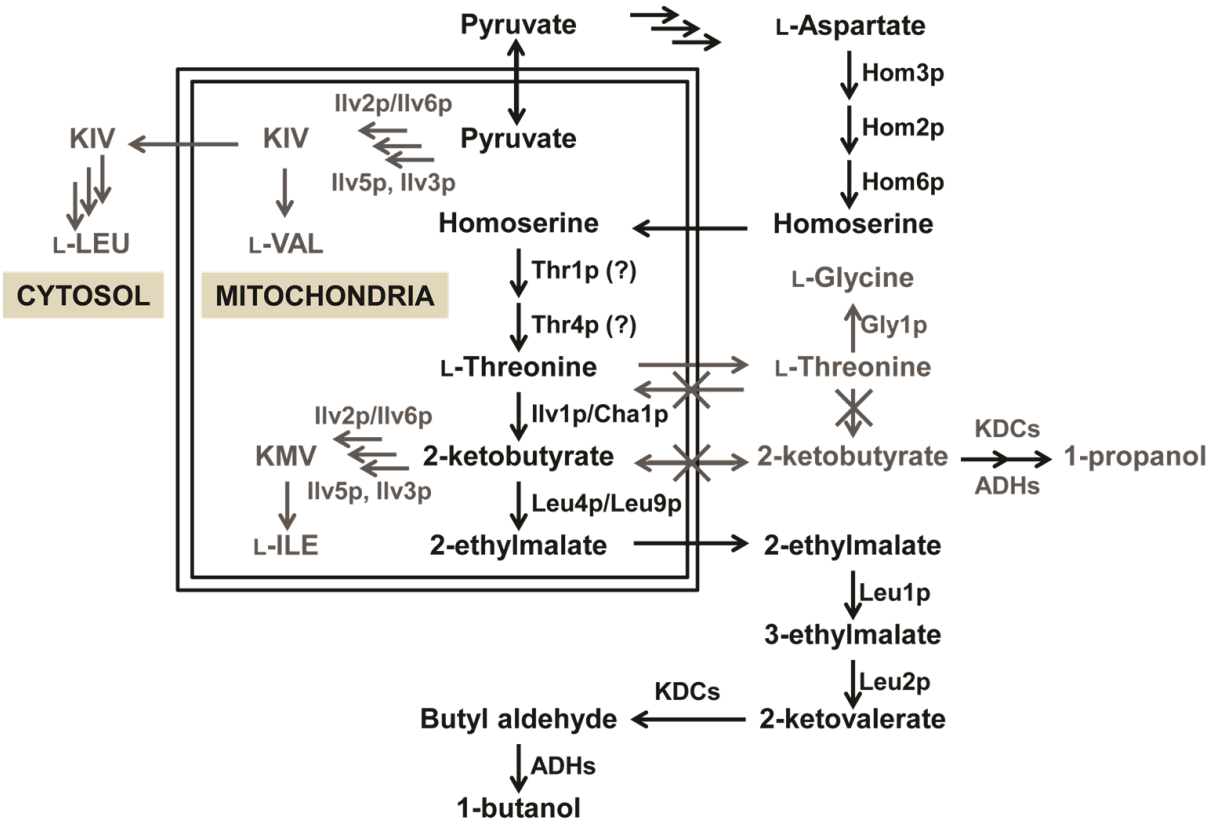

Figure 2.1. Proposed endogenous 1-butanol pathway in *S. cerevisiae.* Co-enzymes and co-factors are omitted for simplicity. Pathways that are not directly involved in 1-butanol production are shown in grey, and the metabolites in these pathways are written in short forms. The information of biochemical pathways and enzyme locations is from literature[23, 24]. The localization of Thr1p and Thr4p are in question, but our results suggest they are mitochondrial enzymes (see **2.3** section and **Table 2.3** for details). KIV: α-ketoisovalerate; KMV: α-keto-β-methylvalerate; LEU: leucine; VAL: valine; ILE: isoleucine; KDCs: 2-keto-acid decarboxylases; ADHs: alcohol dehydrogenases.

In this chapter, we report characterization and engineering of an endogenous 1-butanol pathway in *S. cerevisiae*. Adh1p was found as a switch for the endogenous 1-butanol production, and an amino-acid dependent pathway was proposed which is similar to the metabolic route engineered in *E. coli* (**Fig. 2.1**)[16]. The involvement of the proposed pathway in 1-butanol production was characterized by precursor feeding, ^{13}C-isotope tracer experiments and genetic manipulations. A substantial 1-butanol production level (242.8 mg/L) was achieved in our

metabolically engineered yeast strains, which is the highest 1-butanol titer reported in *S. cerevisiae* from glucose[14].

2.2. Results

2.2.1. *ADH1* as a switch for 1-butanol production

We noted that the wild-type *S. cerevisiae* strain can consume 1-butanol in YPAD medium (**Fig. 2.2A**), which contains ~200 mg/L 1-butanol after filtration and ~70 mg/L 1-butanol after autoclave sterilization. We speculate that alcohol dehydrogenases (ADHs) are responsible for 1-butanol oxidation. To test our hypothesis, the *ADH1* gene, which encoded the most abundant ADH in *S. cerevisiae*[25], was deleted. Adh1p catalyzes the reversible redox reaction between acetaldehyde and ethanol *in vitro*, and the normal function of Adh1p is to reduce acetaldehyde to ethanol employing NADH as a cofactor[25]. In the *adh1*Δ strain, reduced ethanol production (**Fig. 2.2B** and **Fig. 2.3)** and acetaldehyde accumulation (**Fig. 2.11B**) were observed. The *adh1*Δ mutant strain grew poorly compared to the wild type strain (**Fig. 2.3**), probably resulting from acetaldehyde toxicity and disrupted cofactor balance of NAD^+/NADH[26]. Unexpectedly, the *adh1*Δ strain also accumulated 1-butanol under both micro-anaerobic (**Fig. 2.2A**) and aerobic (**Fig. 2.3**) conditions in the YPAD medium.

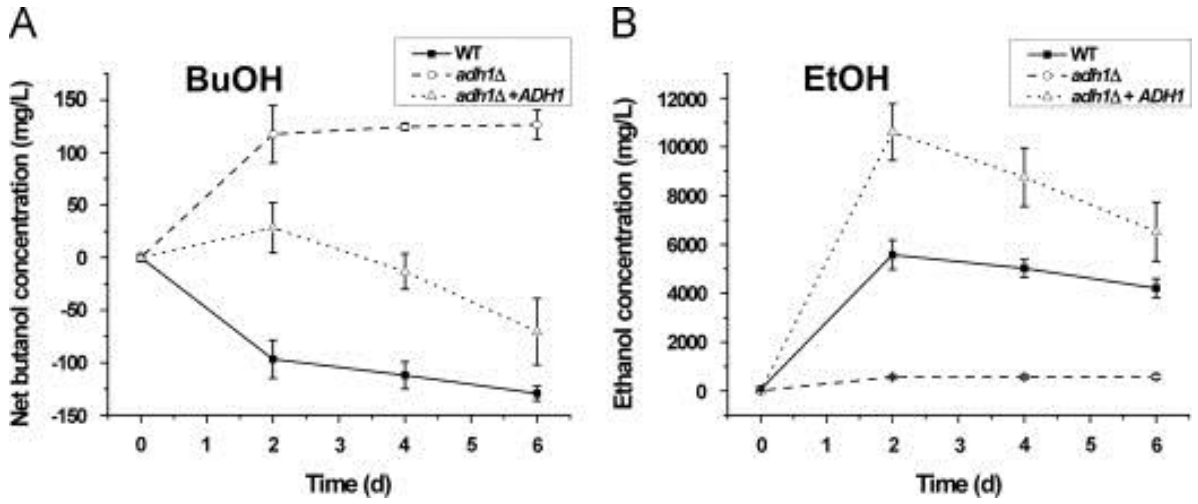

Figure 2.2. Switching on 1-butanol production by *adh1* knockout. Time courses of 1-butanol (*A*) and ethanol (*B*) production in filtered YPAD media under micro-anaerobic condition are shown for the wild type strain (filled square and solid line), the *adh1*Δ strain (open circle and dashed line) and the *adh1*Δ strain with *adh1* overexpressed on pRS425 (open triangle and dotted line). The filtered YPAD medium contains ~200 mg/L 1-butanol at the beginning of fermentation, which is defined as 0 mg/L in the figure. The negative values indicate 1-butanol consumption. Error bars indicate standard deviation of three biological replicates.

Without glucose or the *adh1*Δ strain, no 1-butanol production was observed in the rich medium alone (**Fig. 2.4**). The production of 1-butanol at much lower concentrations by the *adh1*Δ strain was also observed in synthetic dropout medium and citric/phosphate buffer supplemented with only glucose (**Fig. 2.5**). The *adh1*Δ strain with Adh1p over-expression restored 1-butanol consumption at a reduced rate than the wild-type strain (**Fig. 2.2A**), implying that there were other ADHs involved in 1-butanol oxidization. Another reason why the *adh1*Δ deletion can activate 1-butanol production is that when ethanol production is reduced, metabolic flux through the 1-butanol pathway is increased.

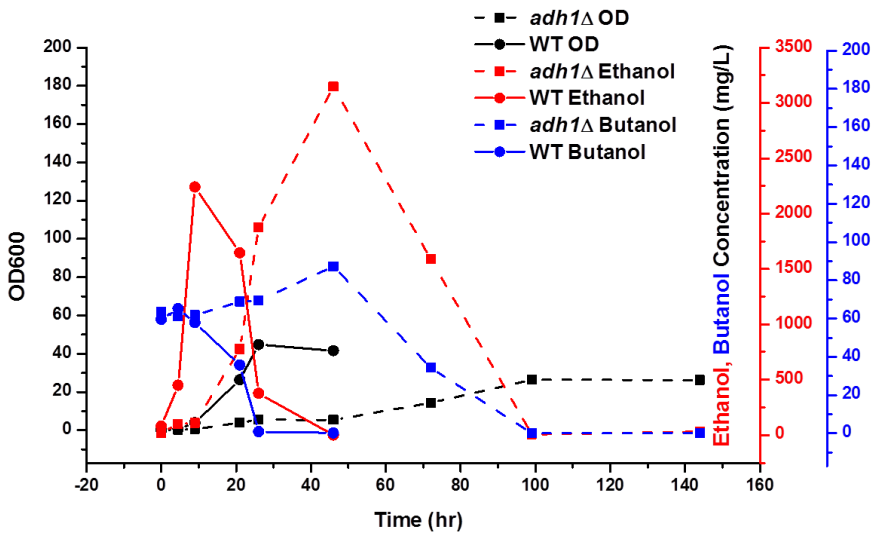

Figure 2.3. Comparison of aerobic fermentation profiles between the wild type strain (solid line) and the *adh1*Δ mutant (dashed line) strain. Fermentation was performed in autoclaved YPAD media (pH=5) under aerobic conditions with a starting OD as 0.1. The experiment has been repeated three times, and one typical set of data is reported.

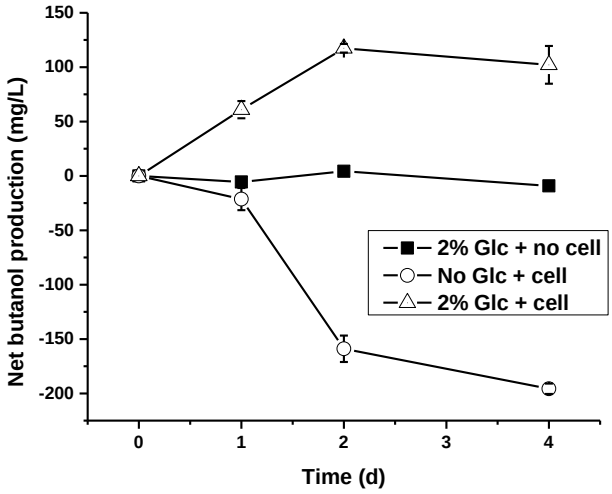

Figure 2.4. No 1-butanol accumulation in the rich medium without glucose or the *adh1*Δ strain. Fermentation was performed in filtered YPA media (pH=5) under micro-anaerobic condition. The filtered YPAD medium contains ~200 mg/L 1-butanol at the beginning of fermentation, which is defined as 0 mg/L in the figure. Error bars indicate standard deviation of three biological replicates.

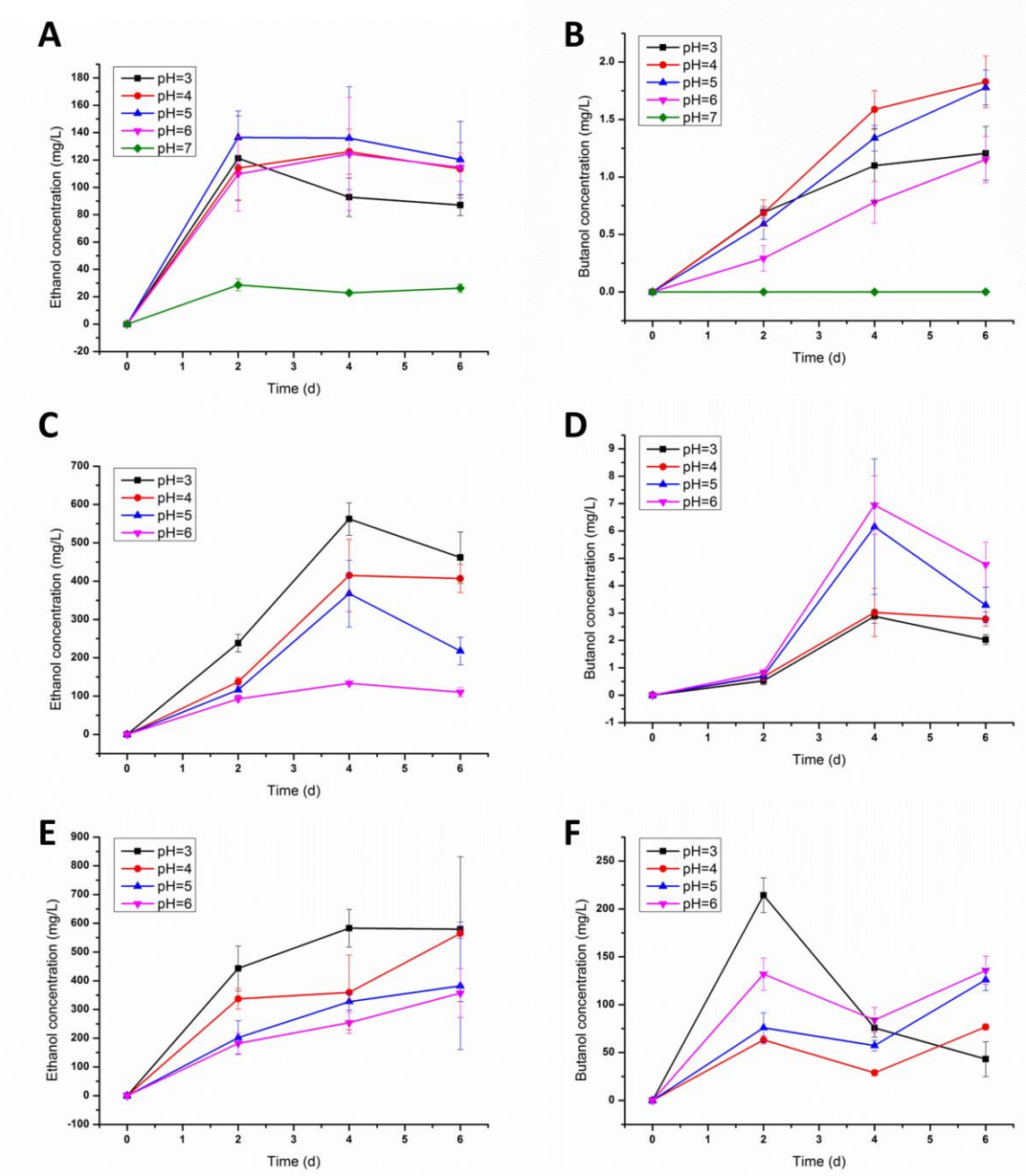

Figure 2.5. Production of 1-butanol by the *adh1*Δ strain from 2% glucose. Three media, citric acid/phosphate buffer (**A** and **B**), SC synthetic dropout medium (**C** and **D**) and filtered YPAD rich medium (**E** and **F**) were tested with different pH values under micro-anaerobic condition. The filtered YPAD medium contains ~200 mg/L 1-butanol at the beginning of fermentation, which is defined as 0 mg/L in the figure. Time courses of ethanol and 1-butanol production are reported in (**A**, **C**, **E**) and (**B**, **D**, **F**), respectively. Error bars indicate standard deviation of three biological replicates.

2.2.2. Precursor feeding studies in the proposed 1-butanol pathway

Based on similarity to the Ehrlich pathway which produces other fusel alcohols[27], we propose an endogenous 1-butanol pathway as follows (**Fig. 2.1**). First, 2-ketobutyrate (2-KB) is generated from L-threonine through a transamination reaction catalyzed by Ilv1p/Cha1p. Then 2-KB is converted to 2-ketovalerate (2-KV) via a keto-acid chain elongation process that is carried out by the leucine biosynthesis pathway enzymes Leu4p/Leu9p, Leu2p and Leu1p[16, 24]. Butyl aldehyde is produced from the decarboxylation of 2-KV by 2-keto acid decarboxylases (KDCs), and is further reduced to 1-butanol by ADHs[21].

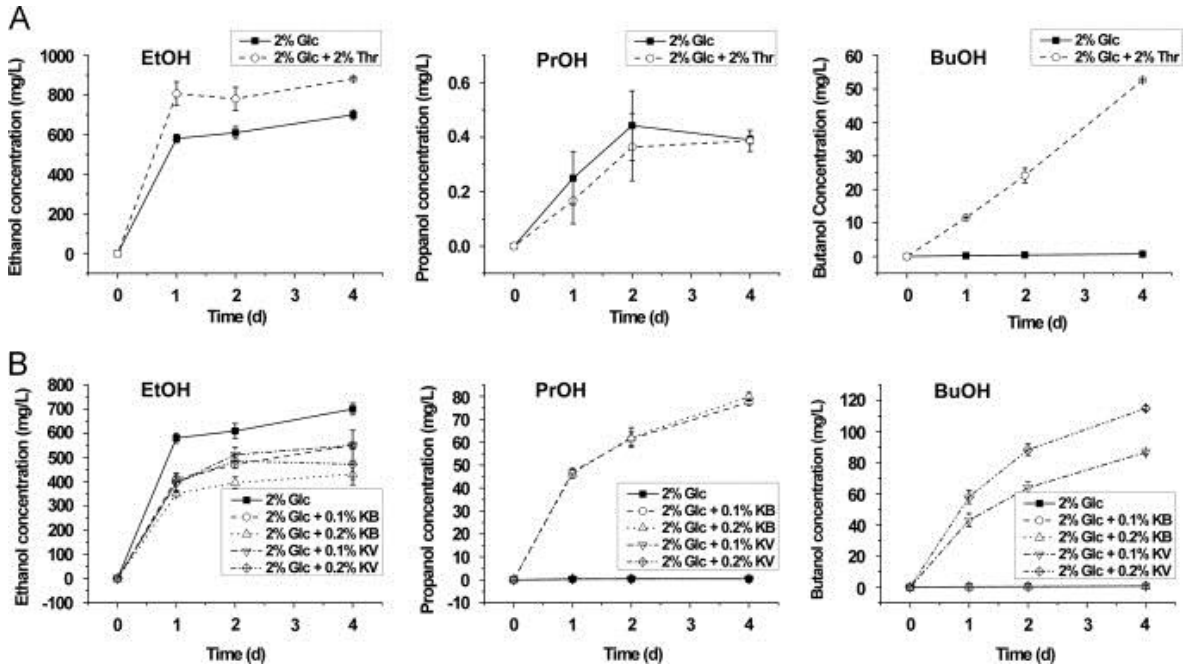

Fig. 2.6. Effects of precursor feeding on alcohol production by the *adh1*Δ strain. Fermentation was performed in citric acid/phosphate buffer with 2% glucose (pH=5) under micro-anaerobic condition with added L-threonine (*A*) or 2-keto acids (*B*). Glc: D-glucose; Thr: L-threonine; KB: 2-ketobutyrate; KV: 2-ketovalerate; EtOH: ethanol; PrOH: 1-propanol; BuOH: 1-butanol. Symbol—2% Glc (closed square), 2% Glc+2% Thr in (*A*) (open circle), 2% Glc+0.1% KB in (*B*) (open circle), 2% Glc+0.2% KB (open triangle), 2% Glc+0.1% KV (open inverse triangle) and 2% Glc+0.2% KV (open diamond). Error bars indicate standard deviation of three biological replicates.

We fed the precursors in the proposed pathway and quantified 1-butanol production. In citric acid/phosphate buffer (pH=5) supplemented with 2% glucose, addition of L-threonine stimulated 1-butanol production in a dose-dependent manner (**Fig. 2.6A** and **Fig. 2.7**).

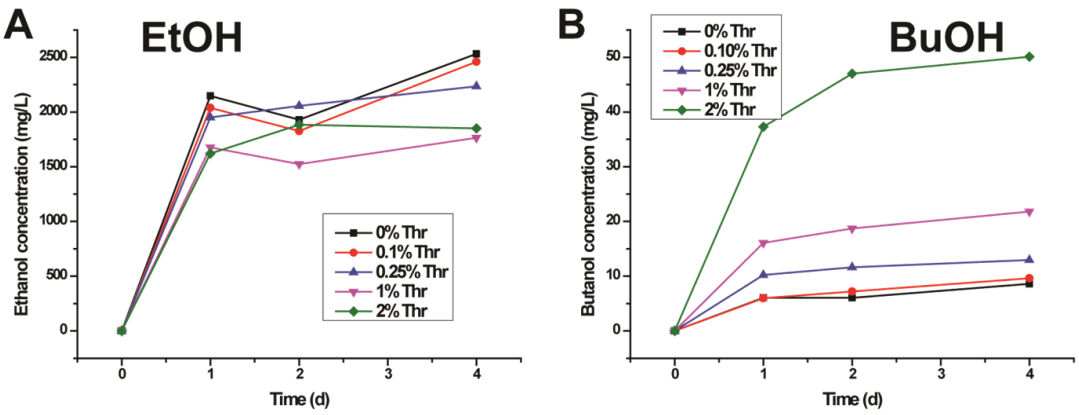

Figure 2.7. Dose-dependent stimulation of 1-butanol production by L-threonine addition. Time courses of ethanol and 1-butanol production are shown in (A) and (B), respectively. Fermentation was performed in citric acid/phosphate buffer with 2% glucose (pH=5) under micro-anaerobic condition. The experiment has been repeated three times, and one representative set of data is reported.

To further elucidate the pathway, ^{13}C-labeled glucose (D-Glucose-$^{13}C_6$) and threonine (L-Threonine-1,2-$^{13}C_2$) were used, and mass spectrometry (MS) fragmentation patterns of the 1-butanol product were analyzed (**Fig. 2.8** and **Fig. 2.9A**). To our surprise, the carbon atoms from exogenous threonine were not incorporated into 1-butanol, and 1-butanol was solely synthesized from glucose (**Fig. 2.9** and **Table 2.1**). Such findings implied that the increased accumulation of 1-butanol was not due to direct transformation of exogenous threonine. Given that threonine deaminase (Ilv1p/Cha1p) and α-isopropylmalate synthase (Leu4p), which catalyze conversion of L-threonine to 2-ethylmalate, are localized in the mitochondria (**Fig. 2.1**)[24], we reason that exogenous threonine cannot be transported from the cytosol into the mitochondria to enter the

proposed 1-butanol pathway.

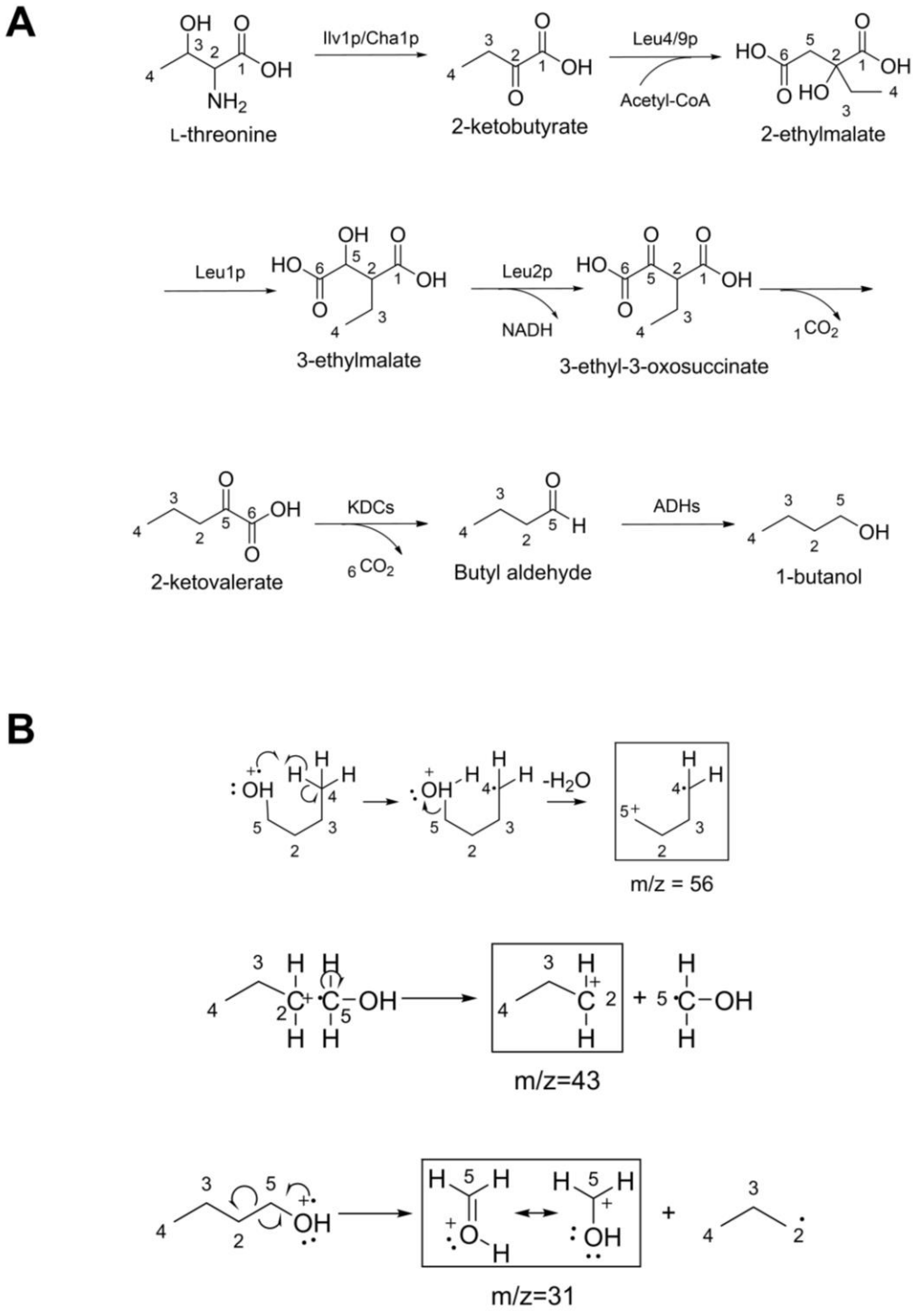

Figure 2.8. Production of 1-butanol from ^{13}C-labeled substrates. (*A*) Biochemical transformation from L-threonine to 1-butanol. (*B*) Generation of fragment ions of 1-butanol in positive electron impact (EI) mode.

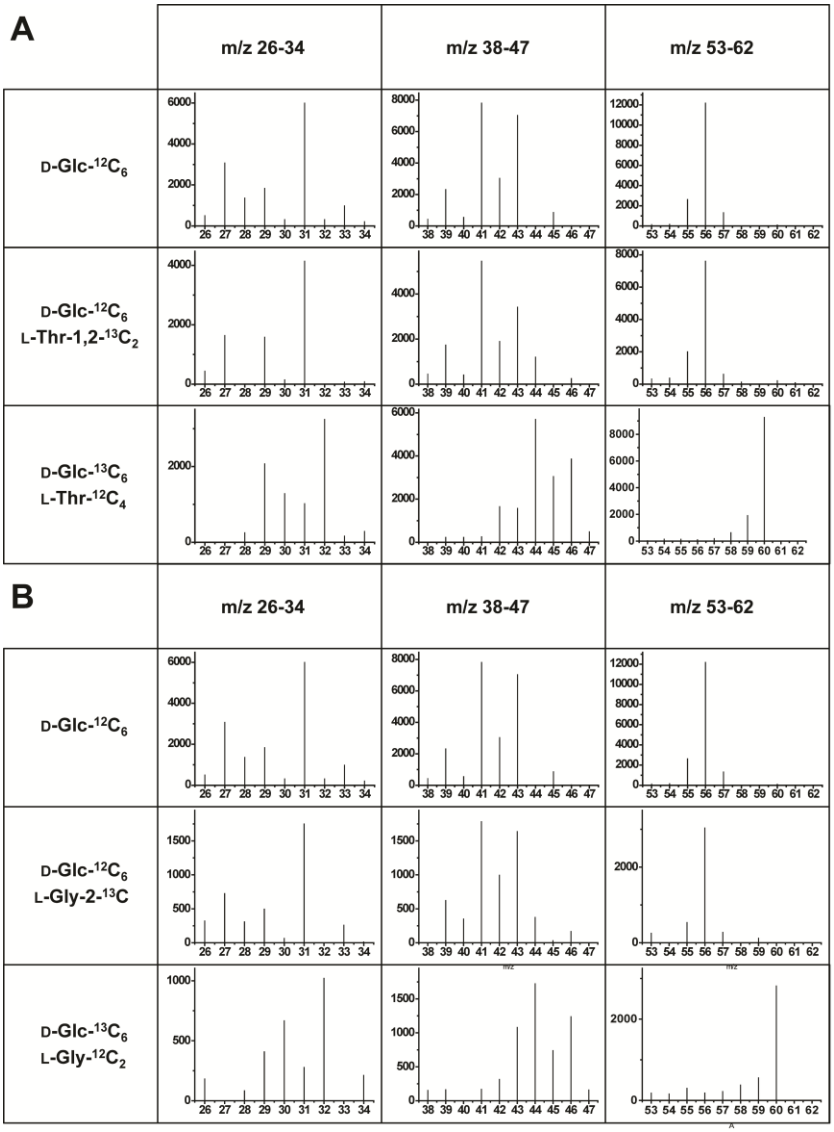

Figure 2.9. Mass spectrum fragmentation patterns of 1-butanol with ^{13}C-labeled substrates. Axis units: *x*-axis: m/z; *y*-axis: arbitrary units. Fermentation was performed in citric acid/phosphate buffer (pH=5) under micro-anaerobic condition, supplemented with 2% glucose and/or 2% amino acids. For 1-butanol produced in medium containing D-Glc-$^{12}C_6$ with or without ^{13}C-labeled amino acids, there were no obvious differences in MS patterns, indicating 1-butanol-$^{12}C_4$ was produced and no ^{13}C atom was incorporated into 1-butanol. For 1-butanol produced in medium containing ^{13}C-labeled glucose and ^{12}C-amino acids, the MS patterns indicated that 1-butanol-$^{13}C_4$ was produced (see the m/z=56 ion) and no ^{12}C atom was incorporated into 1-butanol. The analysis of labeling patterns is summarized in **Table 2.1**.

Table 2.1. Predicted and measured 1-butanol MS fragmentation patterns from ^{13}C-labeled substrates. Fermentation was performed in citric acid/phosphate buffer (pH=5) under micro-anaerobic condition. Measured 1-butanol ^{13}C-labeling patterns match the predicted patterns based on the assumption that 1-butanol is produced from endogenous threonine rather than exogenous threonine. See **Fig. 2.9** for detailed analysis.

	D-Glc-$^{12}C_6$		D-Glc-$^{12}C_6$ L-Thr-1,2-$^{13}C_2$		D-Glc-$^{13}C_6$ L-Thr-$^{12}C_4$	
Labeled Carbon#* in acetyl-CoA	None		None		5, 6	
Carbon source of threonine	Glc	Added	Glc	Added	Glc	Added
Labeled Carbon# in threonine	None	None	None	1,2	1-4	None
Labeled Carbon# in 1-butanol (predicted)	None	None	None	2	2-5	5
Labeled Carbon# in 1-butanol (measured)	None		None**		2-5	

***** Carbon # is referred as in **Fig. 2.8A** and **Fig. 2.8B**

****** Absence of ^{13}C-labeled 1-butanol was also confirmed by ^{13}C NMR (data not shown)

To rationalize how exogenous threonine stimulated 1-butanol production, we hypothesize that it can satisfy cytosolic requirement of threonine for protein expression and glycine synthesis. Thus threonine transport from mitochondria to cytosol is reduced and more threonine is available in the mitochondria for transformation into 1-butanol through the proposed pathway. To test this hypothesis, we also supplemented L-glycine to the medium, suspecting that this would also reduce the requirement for threonine efflux from the mitochondria (**Fig. 2.1**)[28]. We found that L-glycine supplementation indeed resulted in improved 1-butanol production, but to a lesser extent compared to threonine (**Fig. 2.6A** and **Fig. 2.10**). A ^{13}C-labeling study with L-Glycine-2-^{13}C also indicated that exogenous glycine was not directly incorporated into 1-butanol, and all the carbon atoms of 1-butanol were from glucose (**Fig. 2.9**).

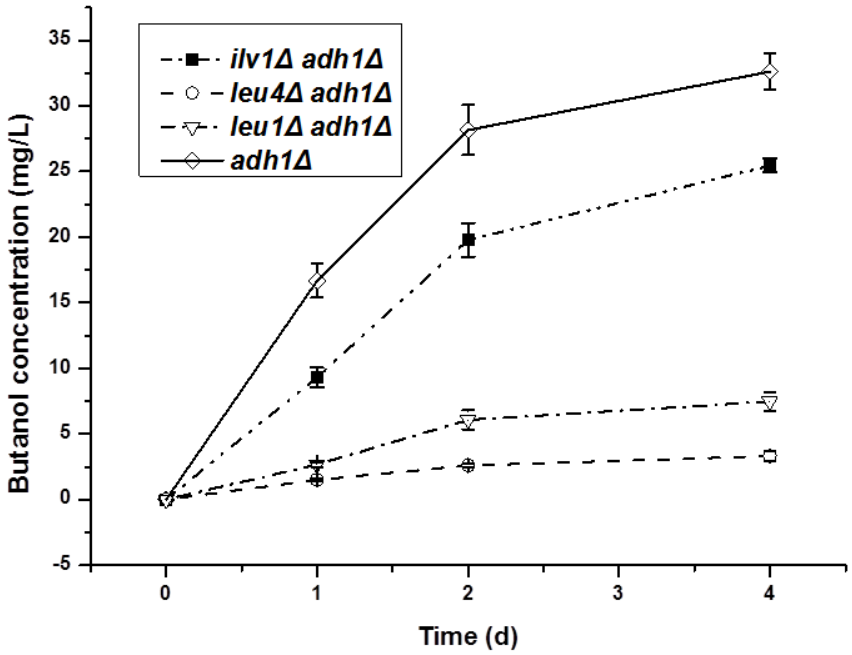

Figure 2.10. Production of 1-butanol with 2% L-glycine addition in knockout strains. Fermentation was performed in citric acid/phosphate buffer with 2% glucose (pH=5) under micro-anaerobic conditions. Error bars indicate standard deviation of three biological replicates.

Addition of 2-KB had no observable effect on 1-butanol production, but resulted in significant 1-propanol production (**Fig. 2.6B** and **Fig. 2.11A**). These results indicate that supplemented 2-KB cannot enter the keto-acid chain elongation process catalyzed by the leucine biosynthesis enzymes. There might be two reasons: 1) 2-KB cannot be transported from the cytosol into the mitochondria, where it is converted to 2-ethylmalate by α-isopropylmalate synthase (**Fig. 2.1**); 2) KDCs have much higher affinity for 2-KB than α-isopropylmalate synthase, and therefore decarboxylation of 2-KB by KDCs prevents incorporation of 2-KB into the 1-butanol pathway. In addition, the very low level of 1-propanol production from glucose with or without added threonine was observed (**Fig. 2.6A**). This phenomenon indicates that the

concentration of cytosolic 2-KB is low, probably due to the lack of cytosolic transamination activity toward threonine for 2-KB synthesis (**Fig. 2.1**).

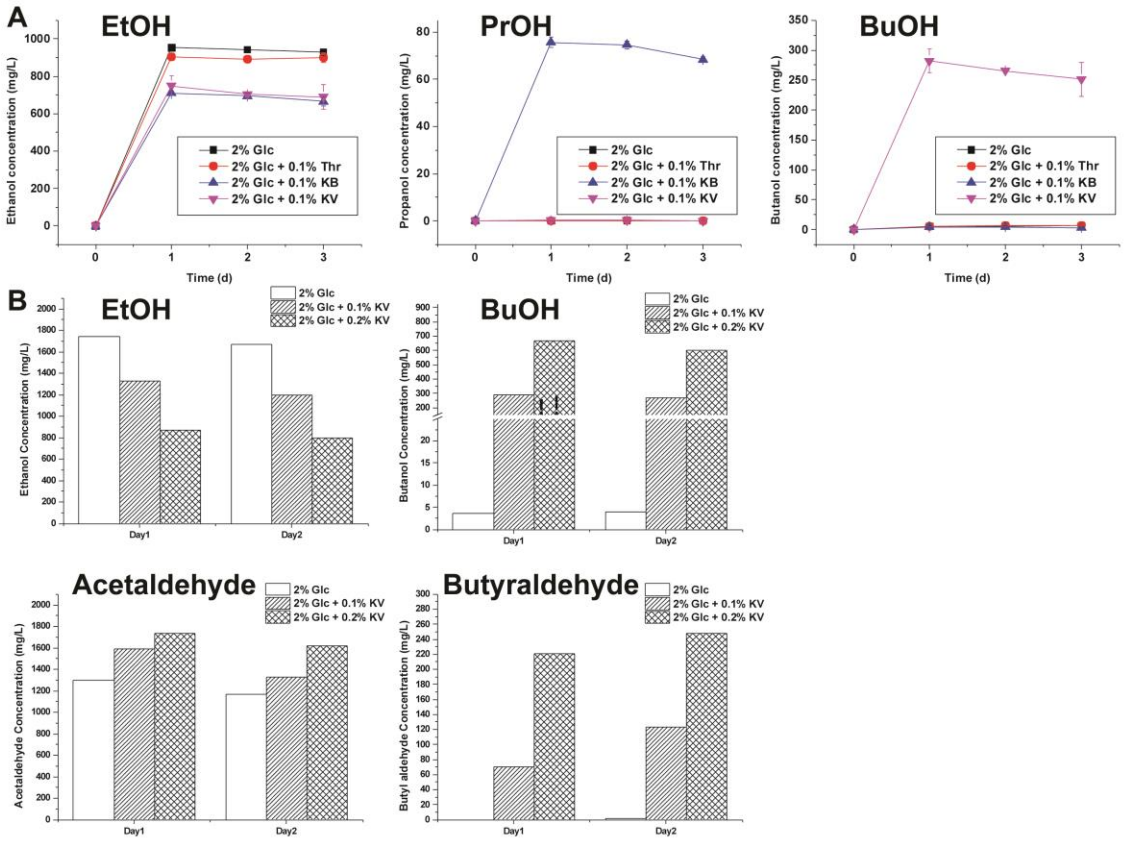

Figure 2.11. Effects of precursor feeding on alcohol production without pH adjustment. (**A**) Time courses of ethanol, 1-propanol and 1-butanol production in the presence of 0.1% added threonine or 2-keto acids in citric acid/phosphate buffer (pH was not adjusted) under micro-anaerobic condition. Symbol type—2% Glc (black square), 2% Glc+0.1% Thr (red circle), 2% Glc+0.1% KB (blue triangle), and 2% Glc+0.1% KV (purple inverse triangle). Error bars indicate standard deviation of two biological replicates. (**B**) Alcohol and aldehyde production with different concentrations of 2-KV addition in citric acid/phosphate buffer (pH was not adjusted) under micro-anaerobic condition. The experiment has been repeated three times, and one representative set of data is reported.

Contrary to threonine and 2-KB, our results indicate that 2-KV can be directly converted to 1-butanol by *S. cerevisiae*. At pH = 5, a condition under which most 2-KV (pK_a=3.38) is in the

dissociated form that cannot freely diffuse across the plasma membrane, only moderate 1-butanol production was observed (0.08 g 1-butanol/g 2-KV, **Fig. 2.6B**). Without pH adjustment, addition of 2-KV greatly lowered the pH value of the medium, and 2-KV should mainly exist in the un-dissociated form. In this case, substantial conversion of 2-KV to 1-butanol and butyl aldehyde was observed (**Fig. 2.11B**, yield of 1-butanol plus butyl aldehyde from 2-KV: 0.36 g/g (0.1% KV), 0.44 g/g (0.2% KV), and 0.64 g/g (theoretical yield)).

2.2.3. Characterization of the 1-butanol pathway by gene deletion

We investigated how deletion of individual genes from the proposed 1-butanol pathway (**Fig. 2.1**) could affect 1-butanol production in the *adh1*Δ background. Due to the greatly reduced fitness of the *adh1*Δ strain, we were unable to obtain the following double-mutant strains: *adh1*Δ/*thr1*Δ, *adh1*Δ/*thr4*Δ, and *adh1*Δ/*cha1*Δ. For the constructed double mutants, *ilv1*Δ or *leu1*Δ led to substantial reduction of 1-butanol production from glucose in citric acid/phosphate buffer (pH=5), and *leu4*Δ nearly abolished 1-butanol accumulation (**Fig. 2.12A**). Deletion of the *LEU9* gene, which encodes an isoenzyme of Leu4p[24], had no obvious effect on 1-butanol production (**Fig. 2.12A**), suggesting that Leu9p plays a minor role in the aldol addition of acetyl-CoA to 2-KB. We then explored how these mutants behaved with supplemented threonine. Production of 1-butanol was reduced for all double-deletion mutant strains compared to the *adh1*Δ single-deletion strain in citric acid/phosphate buffer with 2% glucose and 2% L-threonine (pH=5), and reduction was observed to a greater extent relative to the condition with no added threonine (**Fig. 2.12B**). The gene-knockout mutations of *ilv1*Δ, *leu4*Δ and *leu1*Δ also led to reduced 1-butanol production with 2% supplemented L-glycine relative to the *adh1*Δ single-deletion strain (**Fig. 2.10**). Together, these results indicate that 1-butanol production from

glucose is dependent on the proposed pathway, and the stimulated 1-butanol accumulation by exogenous L-threonine and L-glycine is also related to the proposed pathway.

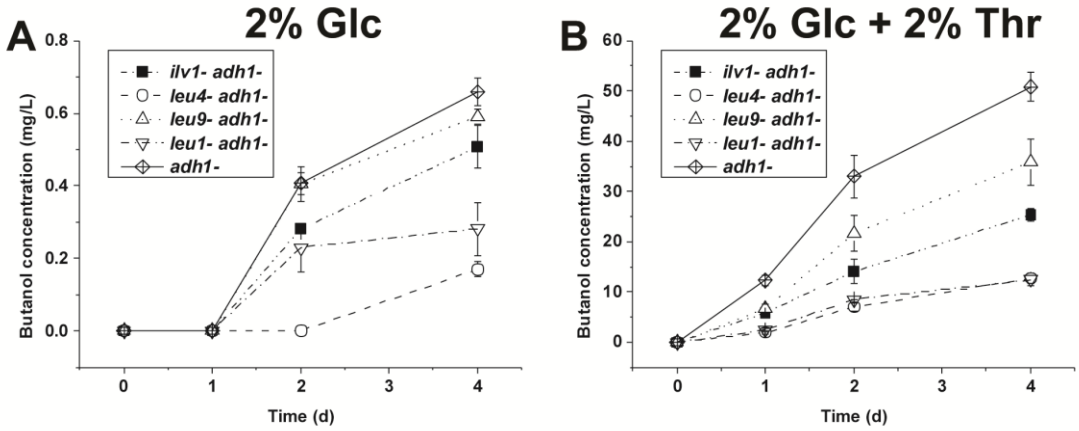

Figure 2.12. Comparison of 1-butanol production in various gene knockout strains. A second gene deletion was introduced in the *adh1*Δ strain. Fermentation was performed in citric acid/phosphate buffer with 2% glucose (pH=5) under micro-anaerobic condition, with (*A*) or without (*B*) 2% L-threonine supplemented. Symbol—*ilv1*Δ *adh1*Δ (closed square), *leu4*Δ *adh1*Δ (open circle), *leu9*Δ *adh1*Δ (open triangle), *leu1*Δ *adh1*Δ (open inverse triangle) and *adh1*Δ (open diamond). Glc: D-glucose; Thr: L-threonine. Error bars indicate standard deviation of three biological replicates.

2.2.4. Gene overexpression to improve 1-butanol production

To re-direct metabolic flux toward the endogenous 1-butanol pathway, the genes in the proposed pathway were over-expressed individually or in combination (**Fig. 2.13**). Resting cells were cultured under micro-aerobic condition in the filtered YPAD medium. Accumulation of 1-butanol decreased after reaching a maximum, probably due to activation of Adh2p which is known for its ability to oxidize alcohols[25]. Thus, comparison of the 1-butanol production was performed with the titer data from day one and day two, when the butanol production peak had not yet been reached (**Fig. 2.13** and **Table 2.2**). Unless indicated otherwise, pRS425, which harbors a *LEU2* expression cassette, was the backbone plasmid for most of the constructs (**Table**

2.2). Background α-isopropylmalate dehydrogenase activity has been reported in laboratory *S. cerevisiae* strains that lack a functional *LEU2* gene[29], and additional copies of the *LEU2* gene on the pRS425 plasmid resulted in only slightly increased 1-butanol production after day two (**Fig. 2.13A**).

To increase threonine biosynthesis, five genes encoding for enzymes that are responsible for converting aspartic acid to threonine (i.e., *hom3*, *hom2*, *hom6*, *thr1* and *thr4*) were over-expressed simultaneously on a multi-copy plasmid pRS426-THR[30]. The strains harboring pRS426-THR accumulated 280.8±77.2% more 1-butanol than the *adh1*Δ strain on the first day, but the 1-butanol concentration did not increase after the first day (**Table 2.2**). To increase 2-KB availability, additional copies of the *ILV1* and *CHA1* genes were introduced on the pRS425 plasmid. On day two, 42.9±21.4% and 73.1±9.2% improvement in 1-butanol production was detected with over-expression of Ilv1p and Cha1p, respectively (**Fig. 2.13A**). Enhanced conversion from 2-KB to 2-KV was achieved by over-expression of the *LEU4* and *LEU1* genes on pRS425, and the strain with the resultant plasmid pRS425-LEU reached a 51.2±20.8% higher 1-butanol titer than the *adh1*Δ strain on day two (**Fig. 2.13A**).

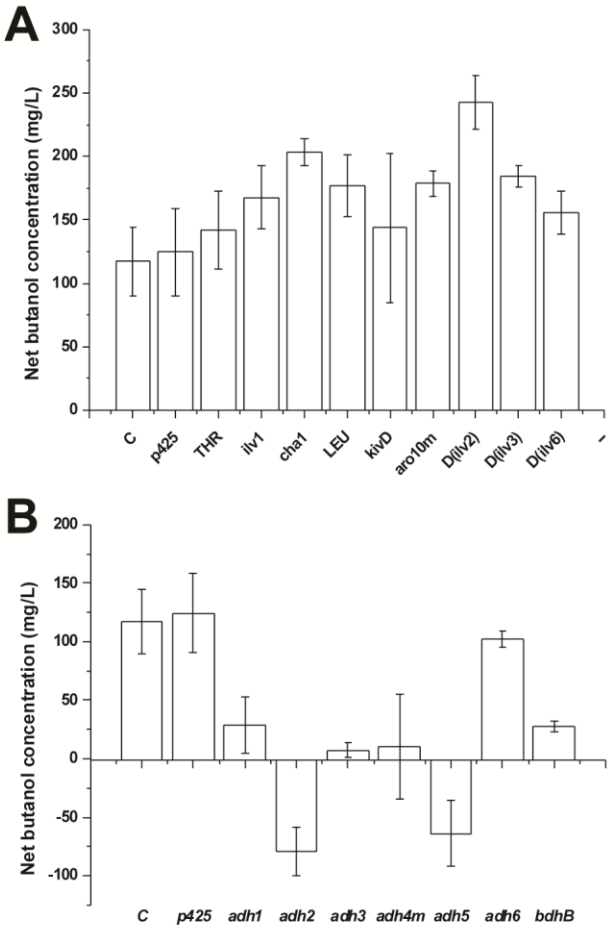

Figure 2.13. 1-Butanol production by metabolically engineered strains. Recombinant strains were constructed in the *adh1*Δ strain (denoted as *C*). p425 is an empty plasmid harboring a *LEU2* expression cassette. For over-expression, backbone vectors are omitted for simplicity and can be found in **Table 2.2**. For double-mutant strains, only the second knockout in addition to *adh1*Δ is shown as D(gene name). Fermentation was performed in filtered YPAD media (pH=5) under micro-anaerobic condition. The filtered YPAD medium contains ~200 mg/L 1-butanol at the beginning of fermentation, which is defined as 0 mg/L in the figure. The net butanol production on Day 2 was reported. All metabolic engineering efforts are summarized in (*A*) except for over-expression of ADHs in (*B*). Error bars indicate standard deviation of three biological replicates.

Regarding KDCs, it has been previously reported that decarboxylation of linear-chain 2-keto acids, pyruvate, 2-KB and 2-KV, is catalyzed exclusively by Pdc1p, Pdc5p, and Pdc6p, and not

by Aro10p or Thi3p[21]. Given the preference of these enzymes toward pyruvate over 2-KV[21], we did not test over-expression of these KDCs, which we speculated would lead to increased production of ethanol rather than 1-butanol. Two other KDCs, KivD from *Lactococcus lactis*[15] and an *S. cerevisiae* Aro10p^{I355Y} mutant[31], were tested on the basis of their reported activity and preference toward 2-KV. Whereas KivD had no significant effect on 1-butanol production, the strain over-expressing the Aro10p^{I355Y} mutant accumulated 52.1±8.5% more 1-butanol than the *adh1*Δ strain on day two (**Fig. 2.13A**).

For the last step, ADHs catalyze reduction of butyl aldehyde to 1-butanol. Six ADHs from *S. cerevisiae* (Adh1p-Adh6p) together with BdhB from *Clostridium acetobutylicum* were over-expressed in the *adh1*Δ background (**Fig. 2.13B**). The mitochondrial signal peptide of Adh4p was removed to create a cytosolic mutant, which increased 1-butanol production by 216.5±55.3% on the first day in the *adh1*Δ background (**Table 2.2**). The Adh6p over-expression strain showed a similar 1-butanol production profile as the *adh1*Δ strain (**Fig. 2.13B**). Strains over-expressing other ADHs all exhibited no accumulation or even consumption of 1-butanol (**Fig. 2.13B** and **Table 2.2**).

Table 2.2. Comparison in 1-butanol production of recombinant yeast strains. Fermentation was performed in filtered YPAD medium (pH=5) under micro-anaerobic condition. The filtered YPAD medium contains ~200 mg/L 1-butanol at the beginning of fermentation, which is defined as 0 mg/L in the table. The negative values reported indicate 1-butanol consumption.

1-Butanol concentration (mg/L)	Day 1	Day 2	Day 4	Percentage difference in titer (Day 1)	Percentage difference in titer (Day 2)
*adh1*Δ (C)	41.2±1.0	117.4±27.2	124.5±3.0	0.0±2.4%	0.0±23.2%
pRS425	98.6±20.2	124.4±34.0	142.0±19.7	139.3±49.0%	6.0±29.0%
pRS426-THR	156.9±31.8	142.1±31.0	174.1±43.7	280.8±77.2%	21.1±26.4%
pRS426-THR-R2	121.7±33.8	44.6±35.4	47.1±37.7	195.4±82.0%	-62.0±30.1%
pRS425-*ilv1*	117.6±14.1	167.7±25.2	185.0±18.5	185.4±34.2%	42.9±21.4%
pRS425-*cha1*	n.m.*	203.2±10.8	142.6±17.3	n.a.**	73.1±9.2%
pRS425-LEU	n.m.	177.4±24.4	195.4±25.2	n.a.	51.2±20.8%
pRS425-*kivD*	n.m.	143.6±58.6	120.1±39.6	n.a.	22.3±49.9%
pRS425-*aro10m*	n.m.	178.5±10.0	97.9±44.1	n.a.	52.1±8.5%
pRS425-*adh1*	n.m.	28.4±24.0	-13.2±16.9	n.a.	-75.8±20.4%
pRS425-*adh2*	-52.0±33.7	-78.8±20.6	-112.8±34.0	-226.2±81.8%	-167.1±17.6%
pRS425-*adh3*	25.4±17.9	7.5±6.5	-23.0±3.3	-38.3±43.4%	-93.6±5.5%
pRS425-*adh4m*	130.4±22.8	10.4±45.0	-13.4±14.5	216.5±55.3%	-91.1±38.3
pRS425-*adh5*	-35.6±14.8	-63.8±28.0	-64.9±14.4	-186.4±35.9%	-154.3±23.8%
pRS423-*adh6*	35.9±21.2	102.1±7.4	90.7±12.6	-12.9±51.5%	-13.0±6.3%
pRS423-*bdhB*	48.7±6.6	27.7±4.3	9.8±0.5	18.2±16.0%	-76.4±3.7%
*ilv2*Δ	n.m.	242.8±21.0	199.9±22.3	n.a.	106.8±17.9%
*ilv3*Δ	98.3±0.7	184.1±8.6	147.0±19.6	138.6±1.7%	56.9±7.3%
*ilv6*Δ	83.9±1.2	155.9±17.0	143.9±3.7	103.6±2.9%	32.8±14.5%

n.m.: not measured

n.a.: not available

Errors indicate standard deviation of three biological replicates

2.2.5. Elimination of competing pathways

Competing reactions were eliminated to avoid loss of precursors in the proposed 1-butanol pathway (**Fig. 2.1**). The benefit of eliminating the acetolactate synthase activity (*ilv6*Δ/*ilv2*Δ) is

threefold: (1) it prevents the conversion of 2-KB to L-isoleucine, and thus increases the availability of 2-KB for 1-butanol synthesis; (2) it raises the concentrations of two key intermediates (aspartic acid and acetyl-CoA) by conserving their common precursor, pyruvate, which is otherwise used for L-leucine and L-valine biosynthesis; (3) it reduces flux through the leucine biosynthesis pathway, which could then take a more active participation in 2-KV synthesis. On day two, a 1-butanol titer of 242.8±21.0 mg/L was observed for the *ilv2*Δ/*adh1*Δ double mutant, which was 106.8±17.9% higher than the *adh1*Δ strain (**Fig. 2.13A**). The *ilv6*Δ mutation had a less obvious impact on 1-butanol accumulation (32.8±14.5% improvement, **Fig. 2.13A** and **Table 2.2**). Knockout of *ilv3* was also performed to decrease the isoleucine, leucine and valine biosynthesis, which led to a 1-butanol titer that was 56.9±7.3% more than the single mutant *adh1*Δ strain (**Fig. 2.13A** and **Table 2.2**).

2.3. Conclusions and Future Prospects

Previous efforts toward reconstituting the *Clostridium* 1-butanol pathway in *S. cerevisiae* met with many challenges, including difficulty in functional expression of bacterial proteins, limited availability of cytosolic acetyl-CoA, and sequestration of the CoA supply to synthesize acyl-CoA intermediates[13, 14]. On the contrary, employment of an endogenous amino-acid pathway for 1-butanol synthesis not only avoids issues related to introduction of a heterologous pathway, but also takes advantage of accumulated knowledge for amino-acid hyper-production. Though a similar pathway has been constructed and engineered in *E. coli*[15, 16], performance of this threonine-dependent pathway should be better in *S. cerevisiae* by avoiding 1-propanol co-production[16]. The absence of 1-propanol as a byproduct is enabled by separation of 2-KB from

cytosolic KDCs by the mitochondrial membrane (**Fig. 2.1**). In addition, computational metabolic simulations suggested that *S. cerevisiae* possesses greater potential in higher alcohol overproduction than *E. coli*, upon introduction of extra flexibility in the yeast central metabolism by incorporating metabolic shortcuts from *E. coli* (such as pyridine nucleotide transhydrogenase, the anaplerotic pathways, the Entner-Doudoroff pathway, and acetyl-CoA synthesis by pyruvate formate-lyase)[32]. Together with the advantages of yeast host over bacterial hosts for alcohol fermentation, we believe that the engineering of the proposed pathway for 1-butanol production in *S. cerevisiae* is promising.

We observed that exogenous L-threonine cannot enter mitochondria to be directly incorporated into 1-butanol synthesis. This observation suggests that some enzymes catalyzing the threonine biosynthesis reactions are located in the mitochondria (**Fig. 2.1**). Especially for the Thr1p and Thr4p enzymes that catalyze the last two steps of threonine synthesis, no direct biochemical evidence is available to determine their subcellular localization (**Table 2.3**). The green fluorescence protein (GFP) has been fused to the carboxyl terminal of Thr1p to assign its localization by imaging, but only ambiguous result was obtained[33]. Though the GFP-fused Thr4p protein was localized to cytoplasm and nucleus[33], computational analysis of the Thr4p primary sequence suggested an 82% probability that Thr4p would be transported into mitochondria[34] (**Table 2.3**). To make it more complicated, the C-terminal fusion of GFP has been found to cause mis-localization of a target protein[33]. Together, no clear localization assignment of Thr1p and Thr4p can be made from literature, and our results imply that they are mitochondrial proteins. Further biochemical investigation by subcellular fractionation will be valuable to determine the Thr1p and Thr4p localization.

Table 2.3. Enzyme localization of the aspartic acid and threonine biosynthesis pathway in *S. cerevisiae*

Enzyme	Predication 1[34]	Predication 2[33]	Other references
Aat1p	0.98*	Mitochondria	Mitochondria[35]
Aat2p	0.05	Cytoplasm	Cytoplasm/Peroxisome[36]
Hom3p	0.20	Cytoplasm	
Hom2p	0.18	Cytoplasm/Nucleus	Plasma membrane[37]
Hom6p	0.14	Cytoplasm/Nucleus	
Thr1p	0.06	Ambiguous	Unknown (www.yeastgenome.org)
Thr4p	0.82	Cytoplasm /Nucleus	

*The value indicates the probability of export to mitochondria, calculated from the MitoProt algorithm[34].

We found that exogenous L-glycine boosted 1-butanol accumulation but cannot be converted directly into 1-butanol (**Fig. 2.9**), and also noted that the stimulation of 1-butanol production by L-glycine was dependent on the proposed pathway (**Fig. 2.10**). These findings provides an alternative explanation to the observations in a recent report that suggested a metabolic route for 1-butanol production with L-glycine and butyl-CoA as substrates in *S. cerevisiae*[29]. Based on that route, ^{13}C atom of L-Glycine-2-^{13}C would be incorporated in 1-butanol, but it was not observed in this study (**Fig. 2.9**). Also, involvement of butyl-CoA in 1-butanol production is questionable, as butyl-CoA concentration is rather low in wild type *S. cerevisiae*[13]. Therefore, we argue that the stimulation of 1-butanol production by L-glycine may be due to the reduced conversion of L-threonine to L-glycine in the presence of exogenous L-glycine, resulting in increased transformation of L-threonine to 1-butanol through our proposed pathway.

The reduced 1-butanol accumulation by gene knockout mutant strains suggested the involvement of the proposed pathway. However, 1-butanol production was not 100% abolished.

There are two possibilities. First, the isoenzymes catalyzing the same reaction (Ilv1p/Cha1p or Leu4p/Leu9p) may account for the residual 1-butanol production. Second, there might also be other endogenous routes for 1-butanol production. Over-expression of the enzymes in the proposed pathway led to increased 1-butanol titer to various degrees. Moderate improvement by over-expression of the wild-type threonine biosynthesis pathway is consistent with the previous observation that threonine biosynthesis is tightly regulated[30]. It is interesting to further investigate whether feedback-inhibition mutant enzymes[30] can enable deregulation and over-production of L-threonine and 1-butanol. Enhanced threonine transamination activity by Ilv1p/Cha1p over-expression led to substantial improvement in 1-butanol titer, indicating this is a key step to direct the metabolic flux toward the proposed pathway. The fact that Cha1p over-expression led to higher 1-butanol accumulation than that of Ilv1p was consistent with the results of the gene knockout experiment, where *ilv1*Δ resulted in only 23.0±9.0% reduction of 1-butanol production (**Fig. 2.12A**). Cha1p might also play a role in the stimulated 1-butanol production by exogenous L-threonine, as Cha1p is transcriptional induced by L-threonine[38, 39]. Over-expression of the leucine biosynthesis pathway resulted in only limited improvement in 1-butanol production, which might be due to the fact that conversion of 2-KB to 2-KV is not a natural biochemical transformation catalyzed by the leucine biosynthesis pathway[15]. The leucine pathway naturally takes 2-ketoisovalerate (KIV) as its substrate (**Fig. 2.1**), which possesses one more methyl group than 2-KB. It is therefore highly possible that the catalytic reactions with 2-KB as the starting precursor is not as efficient as the leucine synthesis. No specific wild-type KDCs or ADHs enzymes were found to catalyze transformation of 2-KV to butyl aldehyde and to 1-butanol. Engineered or heterologous KDCs and ADHs with preferences toward 2-KV and butyl aldehyde, respectively, are therefore highly desirable to selectively pull out the 2-keto acid

intermediate to form 1-butanol. Elimination of the valine, leucine and isoleucine biosynthesis pathway by *ilv2*Δ resulted in the highest 1-butanol production in this study, which is consistent with the observation in *E. coli*[16]. As ethanol was still the main fermentation product by the *adh1*Δ strain, metabolic engineering effort to eliminate ethanol production should greatly improve 1-butanol productivity. Possible strategies include disruption of all ADH genes[40] or creation of a KDC-negative *S. cerevisiae* strain[41]. Particularly, gene deletion of the *ADH2* gene is highly desirable as it oxidizes 1-butanol once being activated (**Fig. 2.3** and **Table 2.2**).

In addition to 1-butanol, other short-chain alcohols have also been proposed as advanced fuel alternatives[4, 42]. These alcohols are naturally found as trace products in yeast fermentation and known as "fusel alcohols"[27]. The Erlich pathway is responsible for the biosynthesis of these alcohols, from 2-keto acids as the degradation intermediates of branched-chain amino acids[27]. Recently, introduction of a broad specificity KDC and the yeast ADH2 has activated the Erlich pathway in *E. coli* and enabled the accumulation of isobutanol, 1-propanol, 1-butanol, 2-methyl-1-butanol, 3-methyl-1-butanol and 2-phenylethanol[15]. Among the fusel alcohols, the branched-chain isomer of 1-butanol, isobutanol, has been the focus of recent studies in both *E. coli* and yeast[4, 42]. In *E. coli*, the enhanced conversion of pyruvate to 2-ketoisovalerate (KIV), together with the decreased competition of pyruvate, resulted in a titer of 22 g/L isobutanol[15]. In *S. cerevisiae*, the engineering has been complicated by the organelle compartmentation, where the synthesis of KIV from L-valine and the conversion of KIV to isobutanol are performed in the mitochondria and the cytosol, respectively[4, 27]. Therefore, by now the most successful strategy to overproduce isobutanol in yeast (0.63 g/L) involves cytosolic over-expression of Ilv2p, Ilv5p, Ilv3p mutants without mitochondria targeting signals, as well as inactivation of the endogenous mitochondrial pathway by deleting the first enzyme Ilv2p[43]. In addition, highly specific KDC

and ADH enzymes have also been screened to further improve isobutanol production[43, 44]. These strategies may offer useful perspective on the engineering of the endogenous 1-butanol pathway in the present study, given its similarities with the isobutanol pathway. Recent research on the isobutanol biosynthesis engineering in *S. cerevisiae* is summarized in an excellent review[4].

In summary, we reported the endogenous 1-butanol production in *S. cerevisiae* from a renewable substrate, glucose, upon deletion of the ADH1 gene. We proposed and characterized an endogenous 1-butanol pathway in S. *cerevisiae*, and showed potential for over-production of 1-butanol by metabolic engineering. Combination of beneficial manipulations and systematic strain engineering are in progress to further improve the titer and productivity of 1-butanol fermentation. Given the accumulated knowledge about amino-acid biosynthesis and regulation, it is believed that high productivity of 1-butanol production can be achieved through this threonine-dependent pathway by further strain engineering.

2.4. Materials and methods

2.4.1. Media and cultivation conditions

Citric acid/phosphate buffer was used to prepare culture media with controlled pH (http://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers/). *S. cerevisiae* strains were cultivated in either synthetic dropout medium (0.17% Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate and 0.083% amino acid drop out mix, 0.01% adenine hemisulfate and 2% glucose) or YPAD medium (1% yeast extract, 2% peptone, 0.01% adenine hemisulfate and 2% glucose). For 1-butanol production in precursor feeding and gene knockout experiments, only 2% glucose and indicated precursors were added

into citric acid/phosphate buffer. Media pH was adjusted to 5 after addition of precursors by 12 M HCl or 10 M NaOH. To select and maintain strains carrying the KanMX marker, G418 (200 mg/L) was supplemented in YPAD medium. *S. cerevisiae* strains were cultured at 30 °C with 250 rpm agitation in baffled shake-flasks for aerobic growth. For micro-anaerobic fermentation, 4 mL cultures were grown at 30 °C and 250 rpm in Bellco 18×150 mm anaerobic glass tubes sealed with rubber stoppers and aluminum crimps (Chemglass, Vineland, NJ). Vacuum was applied through a syringe needle for 20 minutes, and sterile nitrogen was then added to create micro-aerobic conditions. *E. coli* strains were cultured at 37 °C and 250 rpm in Luria broth (LB) medium (Fisher Scientific, Pittsburgh, PA) with the supplement of 100 μg/mL ampicillin. All chemicals were purchased from Sigma-Aldrich or Fisher Scientific.

2.4.2. DNA manipulation

Plasmid cloning was performed by In-fusion HD cloning Kit (Clontech Laboratories, Mountain View, CA) following the manufacturer's instructions or by the DNA assembler method[45]. The complete list of plasmids in this study is summarized in **Table 2.4**. For DNA manipulations, yeast plasmids were isolated using a Zymoprep II yeast plasmid isolation kit (Zymo Research, Irvine, CA) and transferred into *E. coli* for amplification. QIAprep Spin Plasmid Mini-prep Kits (Qiagen, Valencia, CA) were employed to prepare plasmid DNA from *E. coli*. Yeast genomic DNA was isolated by Wizard Genomic DNA Purification Kit (Promega, Madison, WI). All enzymes used for recombinant DNA cloning were from New England Biolabs (Ipswich, MA) unless otherwise noted. The products of PCR, digestion and ligation reactions were purified by QIAquick PCR Purification and Gel Extraction Kits (Qiagen, Valencia, CA).

2.4.3. Strain constructions

The complete list of strains in this study was summarized in **Table 2.5**. *S. cerevisiae* strain YSG50 *ura3*Δ (*MATα ura3*Δ *ade2-1 his3-11,15 leu2-3,112 can1-100 trp1-1*) was designated as the wild type (WT) strain[46]. Zymo 5α Z-competent *E. coli* (Zymo Research, Irvine, CA) was used for yeast plasmid amplification. To perform gene deletions, the open reading frames (ORF) of target genes were disrupted by inserting the LoxP-*ura3*-LoxP or LoxP-KanMX-LoxP cassettes which were PCR-amplified from plasmid pUG72 or pUG6, respectively[47]. Between consecutive gene deletions, the selection marker was recycled through the Cre-LoxP recombination mechanism by pSH47 or pSH63[47]. Pathway integration into the delta sites in the yeast genome was achieved by a home-made integration plasmid. Following *Pme*I digestion, the recombinant plasmid would release a linear DNA fragment in which the pathway to integrate and the KanMX marker were flanked by delta sequences.

2.4.4. Metabolite detection

Alcohol and aldehyde compounds were quantified by an Agilent 7890 gas chromatograph equipped with an Agilent 5975 mass selective detector (Agilent Inc., Palo Alto, CA) at the Roy J. Carver Metabolomics Center (University of Illinois, Urbana, IL). Filtered culture supernatant was stored at -80 ºC prior to analysis to minimize evaporation. Samples (1 μL) were injected in split mode (10:1), and analyzed on a 15 m Zebron ZB-FFAP column with 0.25 mm I.D. and 0.25 μm film thickness (Phenomenex, Torrance, CA). Injection port and interface temperature was set at 250 ºC, and the ion source set to 230 ºC. The helium carrier gas was set at a constant flow rate of 1.6 mL/min. The oven temperature program was set as the following: a) 2 min isothermal

heating at 60 ºC, b) increase at a rate of 10 ºC min^{-1} to 80 ºC for 0 min, c) and then 40 ºC min^{-1} to 200 ºC for the final 4 min. The mass spectrometer was operated in positive electron impact mode (EI) at 69.9 eV ionization energy with an m/z 25-300 scan range. All chromatogram spectra were analyzed with the HP Chemstation program (Agilent, Palo Alto, CA, USA).

Table 2.4. List of plasmids in this study.

Name	Description
pRS425	2μ *ori*, LEU2 marker
pRS423	2μ *ori*, HIS3 marker
pRS426	2μ *ori*, URA3 marker
pRS426-THR	P_{TEF1}-*hom3*-T_{PGK1}-P_{TPI1}-*hom2*-T_{GPD1}-P_{ENO2}-*hom6*-T_{TEF1}-P_{PDC1}-*thr1*-T_{HXT7}-P_{FBA1}-*thr4*-T_{TEF2}
pRS425-*ilv1*	P_{TEF1}-*ilv1*-T_{PGK1}
pRS425-*cha1*	P_{TEF1}-*cha1*-T_{PGK1}
pRS425-LEU	P_{TEF1}-*leu4*-T_{PGK1}-P_{TPI1}-*leu1*-T_{GPD1}
pRS425-*kivD*	P_{TEF1}-*kivD*-T_{PGK1}
pRS425-*aro10m*	P_{TEF1}-*aro10m*-T_{PGK1} (*aro10m* encodes Aro10p^{I355Y})
pRS425-*adh1*	P_{TEF1}-*adh1*-T_{PGK1}
pRS425-*adh2*	P_{TEF1}-*adh2*-T_{PGK1}
pRS425-*adh3*	P_{TEF1}-*adh3*-T_{PGK1}
pRS425-*adh4m*	P_{TEF1}-*adh4m*-T_{PGK1} (*adh4m* as a mutant of *adh4* with first 129 nt truncated)
pRS425-*adh5*	P_{TEF1}-*adh5*-T_{PGK1}
pRS423-*adh6*	P_{TEF1}-*adh6*-T_{PGK1}
pRS423-*bdhB*	P_{TEF1}-*BdhB*-T_{PGK1}

Table 2.5. List of strains in this study.

Name	Description
YSG50 *ura3*Δ	*MATα ura3*Δ *ade2-1 his3-11,15 leu2-3,112 can1-100 trp1-1*[46].
*adh1*Δ	YSG50 *ura3*Δ *adh1*Δ (with URA3 marker recycled)
*adh1*Δ *ilv1*Δ	YSG50 *ura3*Δ *adh1*Δ *ilv1*Δ::*ura3*
*adh1*Δ *leu4*Δ	YSG50 *ura3*Δ *adh1*Δ *leu4*Δ::*ura3*
*adh1*Δ *leu9*Δ	YSG50 *ura3*Δ *adh1*Δ *leu9*Δ::*ura3*
*adh1*Δ *leu1*Δ	YSG50 *ura3*Δ *adh1*Δ *leu1*Δ::*ura3*
*adh1*Δ *ilv2*Δ	YSG50 *ura3*Δ *adh1*Δ *ilv2*Δ::*ura3*
*adh1*Δ *ilv6*Δ	YSG50 *ura3*Δ *adh1*Δ *ilv6*Δ::*ura3*
*adh1*Δ *ilv3*Δ	YSG50 *ura3*Δ *adh1*Δ *ilv3*Δ::*ura3*

2.5. References

1. Lee, J.W. et al. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol* **8**, 536-546 (2012).
2. Du, J., Shao, Z. & Zhao, H. Engineering microbial factories for synthesis of value-added products. *J Ind Microbiol Biotechnol* **38**, 873-890 (2011).
3. Hong, K.-K. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol Life Sci* **69**, 2671-2690 (2012).
4. Buijs, N.A., Siewers, V. & Nielsen, J. Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Curr Opin Chem Biol* **17**, 480-488 (2013).
5. Peralta-Yahya, P.P., Zhang, F.Z., del Cardayre, S.B. & Keasling, J.D. Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320-328 (2012).
6. Jin, C., Yao, M.F., Liu, H.F., Lee, C.F.F. & Ji, J. Progress in the production and application of n-butanol as a biofuel. *Renew Sust Energ Rev* **15**, 4080-4106 (2011).
7. Dürre, P. Biobutanol: An attractive biofuel. *Biotechnol J* **2**, 1525-1534 (2007).
8. Lee, S.Y. et al. Fermentative butanol production by *Clostridia*. *Biotechnol Bioeng* **101**, 209-228 (2008).
9. Inui, M. et al. Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Appl Microbiol Biotechnol* **77**, 1305-1316 (2008).
10. Atsumi, S. et al. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab Eng* **10**, 305-311 (2008).
11. Dellomonaco, C., Clomburg, J.M., Miller, E.N. & Gonzalez, R. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**, 355-U131 (2011).

12. Shen, C.R. et al. Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in *Escherichia coli*. *Appl Environ Microbiol* **77**, 2905-2915 (2011).

13. Steen, E.J. et al. Metabolic engineering of *Saccharomyces cerevisiae* for the production of *n*-butanol. *Microb Cell Fact* **7**, 8 (2008).

14. Krivoruchko, A., Serrano-Amatriain, C., Chen, Y., Siewers, V. & Nielsen, J. Improving biobutanol production in engineered *Saccharomyces cerevisiae* by manipulation of acetyl-CoA metabolism. *J Ind Microbiol Biotechnol*, 1-6 (2013).

15. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89 (2008).

16. Shen, C.R. & Liao, J.C. Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab Eng* **10**, 312-320 (2008).

17. Knoshaug, E.P. & Zhang, M. Butanol Tolerance in a Selection of Microorganisms. *Appl Biochem Biotechnol* **153**, 13-20 (2009).

18. Gonzalez-Ramos, D., van den Broek, M., van Maris, A.J.A., Pronk, J.T. & Daran, J.M.G. Genome-scale analyses of butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation. *Biotechnol Biofuels* **6**, 48 (2013).

19. Williams, T.C., Nielsen, L.K. & Vickers, C.E. Engineered quorum sensing using pheromone-mediated cell-to-cell communication in *Saccharomyces cerevisiae*. *ACS Synth Biol* **2**, 136-149 (2013).

20. Webb, A.D., Kepner, R.E. & Ikeda, R.M. Composition of Typical Grape Brandy Fusel Oil. *Anal Chem* **24**, 1944-1949 (1952).

21. Romagnoli, G., Luttik, M.A.H., Kotter, P., Pronk, J.T. & Daran, J.M. Substrate Specificity of Thiamine Pyrophosphate-Dependent 2-Oxo-Acid Decarboxylases in

Saccharomyces cerevisiae. *Appl Environ Microbiol* **78**, 7538-7548 (2012).

22. Ingraham, J.L., Guymon, J.F. & Crowell, E.A. The pathway of formation of *n*-butyl and *n*-amyl alcohols by a mutant strain of *Saccharomyces cerevisiae*. *Arch Biochem Biophys* **95**, 169-175 (1961).

23. Strathern, J.N., Jones, E.W. & Broach, J.R. in The molecular biology of the yeast saccharomyces metabolism and gene expression (Cold Spring Harbor laboratory Press, Plainview, New York; 1982).

24. Kohlhaw, G.B. Leucine biosynthesis in fungi: entering metabolism through the back door. *Microbiol Mol Biol Rev* **67**, 1-15 (2003).

25. Leskovac, V., Trivic, S. & Pericin, D. The three zinc-containing alcohol dehydrogenases from baker's yeast, *Saccharomyces cerevisiae*. *FEMS Yeast Res* **2**, 481-494 (2002).

26. Drewke, C., Thielen, J. & Ciriacy, M. Ethanol formation in adh0 mutants reveals the existence of a novel acetaldehyde-reducing activity in *Saccharomyces cerevisiae*. *J Bacteriol* **172**, 3909-3917 (1990).

27. Hazelwood, L.A., Daran, J.M., van Maris, A.J., Pronk, J.T. & Dickinson, J.R. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl Environ Microbiol* **74**, 2259-2266 (2008).

28. Monschau, N., Stahmann, K.P., Sahm, H., McNeil, J.B. & Bognar, A.L. Identification of *Saccharomyces cerevisiae* GLY1 as a threonine aldolase: a key enzyme in glycine biosynthesis. *FEMS Microbiol Lett* **150**, 55-60 (1997).

29. Branduardi, P., Longo, V., Berterame, N.M., Rossi, G. & Porro, D. A novel pathway to produce butanol and isobutanol in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **6**, 68 (2013).

30. Maria-Jose Farfan, I.L.C. Enrichment of threonine content in *Saccharomyces cerevisiae* by pathway engineering. *Enzyme Microb Technol* **26**, 763-770 (2000).

31. Kneen, M.M. et al. Characterization of a thiamin diphosphate-dependent phenylpyruvate decarboxylase from *Saccharomyces cerevisiae*. *FEBS J* **278**, 1842-1853 (2011).

32. Matsuda, F. et al. Engineering strategy of yeast metabolism for higher alcohol production. *Microb Cell Fact* **10**, 70 (2011).

33. Huh, W.K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691 (2003).

34. Claros, M.G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241**, 779-786 (1996).

35. Morin, P.J., Subramanian, G.S. & Gilmore, T.D. Aat1, a gene encoding a mitochondrial aspartate-aminotransferase in *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **1171**, 211-214 (1992).

36. Verleur, N., Elgersma, Y., VanRoermund, C.W.T., Tabak, H.F. & Wanders, R.J.A. Cytosolic aspartate aminotransferase encoded by the AAT2 gene is targeted to the peroxisomes in oleate-grown *Saccharomyces cerevisiae*. *Eur J Biochem* **247**, 972-980 (1997).

37. Delom, F. et al. The plasma membrane proteome of *Saccharomyces cerevisiae* and its response to the antifungal calcofluor. *Proteomics* **6**, 3029-3039 (2006).

38. Pedersen, J.O. et al. Locus-specific suppression of ilv1 in *Saccharomyces cerevisiae* by deregulation of CHA1 transcription. *Mol Gen Genet* **255**, 561-569 (1997).

39. Bornaes, C., Ignjatovic, M.W., Schjerling, P., Kiellandbrandt, M.C. & Holmberg, S. A Regulatory Element in the Cha1 Promoter Which Confers Inducibility by Serine and

Threonine on *Saccharomyces cerevisiae* Genes. *Mol Cell Biol* **13**, 7604-7611 (1993).

40. Ida, Y., Furusawa, C., Hirasawa, T. & Shimizu, H. Stable disruption of ethanol production by deletion of the genes encoding alcohol dehydrogenase isozymes in *Saccharomyces cerevisiae*. *J Biosci Bioeng* **113**, 192-195 (2012).

41. van Maris, A.J.A. et al. Directed evolution of pyruvate decarboxylase-negative *Saccharomyces cerevisiae*, yielding a C-2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl Environ Microbiol* **70**, 159-166 (2004).

42. Wen, M., Bond-Watts, B.B. & Chang, M.C.Y. Production of advanced biofuels in engineered *E. coli*. *Curr Opin Chem Biol* **17**, 472-479 (2013).

43. Brat, D., Weber, C., Lorenzen, W., Bode, H. & Boles, E. Cytosolic re-localization and optimization of valine synthesis and catabolism enables increased isobutanol production with the yeast *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **5**, 65 (2012).

44. Kondo, T. et al. Genetic engineering to enhance the Ehrlich pathway and alter carbon flux for increased isobutanol production from glucose by *Saccharomyces cerevisiae*. *J Biotechnol* **159**, 32-37 (2012).

45. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**, e16 (2009).

46. Nair, N.U. & Zhao, H. Mutagenic inverted repeat assisted genome engineering (MIRAGE). *Nucleic Acids Res* **37**, e9 (2009).

47. Gueldener, U., Heinisch, J., Koehler, G.J., Voss, D. & Hegemann, J.H. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res* **30**, e23 (2002).

Chapter 3 RAGE1.0: Complex Phenotype Engineering

3.1. Introduction

Complex phenotypes, such as inhibiter tolerance, involves synergistic actions of many genes[1]. Such complex phenotypes are often poorly understood and extremely difficult to engineer[2, 3]. Adaptive engineering has been the method-of-choice to isolate evolved strains with improved inhibitor tolerance, through serial transfers with increasing inhibitor stresses in the medium[4]. Though effective, adaptive engineering is very time-consuming because the appearance of mutations is infrequent and most of these mutations are detrimental or neutral[4]. Therefore, new methods are needed to efficiently generate multiplex genetic diversity on a genome scale, as the engineering of complex traits often requires simultaneous modulation of many genes[2, 3, 5, 6].

The current ability to engineer a genome in multiplex is mostly limited to bacterial hosts[3]. A microbial genome can now be synthesized *de novo* which should in principle enable the ultimate genome-scale engineering, but this strategy is limited to small bacterial genomes and is also too expensive and tedious for most genome engineering applications[7]. In *Escherichia coli*, recombination-based genetic engineering (recombineering) enables generation of combinatorial genomic diversity[5] or genome-wide identification of gene targets for a certain trait[8]. However, lack of efficient tools for large-scale DNA oligonucleotide-mediated allelic replacement hinders the application of recombineering in eukaryotes.

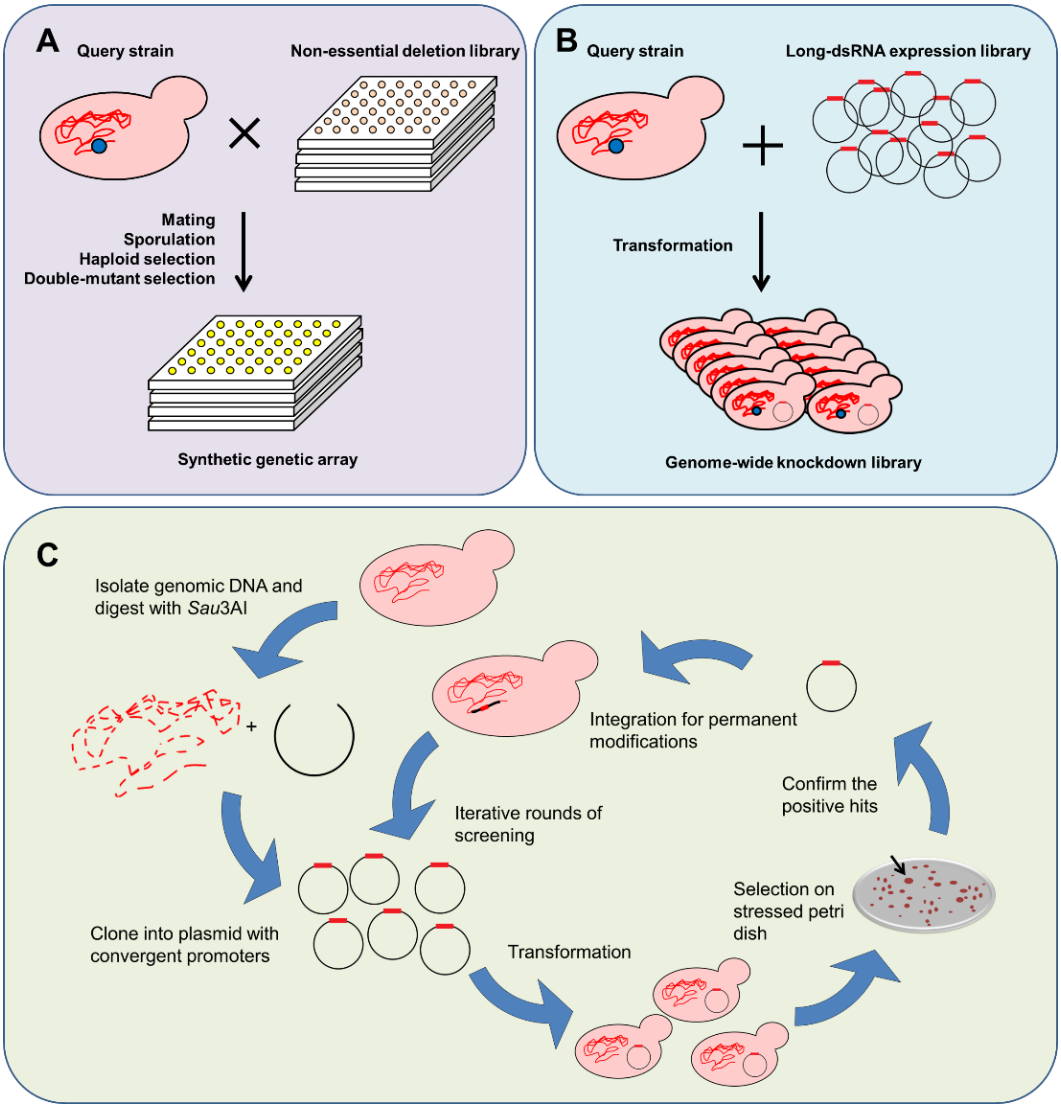

Figure 3.1. RNAi-assisted genome evolution enables rapid cellular reprogramming through iterative rounds of RNAi library creation and high throughput screening or selection. (**A**) Traditional strain libraries approach. Individual gene-knockout is performed to construct a mutant library. Synthetic Genetic Array (SGA) creates an array of double-mutant strains through multiple steps of manipulation. (**B**) Creation of a pooled RNAi library only requires a simple step of transformation. (**C**) The ease of RNAi library construction enables repeated rounds of screening in an evolving genetic background (RAGE), which accumulates the beneficial modifications identified from the previous rounds of screening by integration.

S. cerevisiae is not only a prominent model eukaryotic organism but also a widely used platform organism for industrial production of chemicals and fuels[9, 10]. For this well-studied eukaryote *S. cerevisiae,* non-essential genes have been individually deleted to construct strain libraries for functional screening[11, 12] (**Fig. 3.1A**). Synthetic Genetic Array (SGA) has been developed to assay genetic interactions, whereby a query strain with a modified genetic background is crossed with a gene-deletion library to create an ordered array of haploid double-mutant strains[13] (**Fig. 3.1A**). Although strain libraries have provided invaluable knowledge about numerous important biological processes[14], the tedious procedure to introduce genome-wide perturbations on a wild-type or mutated genome severely limits our ability to reprogram eukaryotic cells (**Fig. 3.1A**). In addition, gene-knockout libraries are only available for certain laboratory strains of the *S. cerevisiae* species[15], whereas different strains exhibit dramatic differences in phenotypes[16]. Therefore, effective tools to iteratively introduce genome-wide modifications in customized genetic backgrounds are highly desirable for successful genome engineering practice in yeast.

Here we report RAGE for engineering complex traits in yeast by directed genome evolution with iterative RNAi screening. Directed evolution mimics Darwinian evolution in a test tube and involves iterative rounds of genetic diversification and high throughput screening or selection, and has achieved enormous success in tailoring biological systems ranging from single proteins to whole cells[17, 18]. However, no satisfying tools exist to apply directed evolution strategy on a genome scale in *S. cerevisiae*, as the current method to introduce genome-wide mutations in an evolving yeast genome is prohibitively tedious (**Fig. 3.1A**). On the other hand, the introduction of a pooled RNAi library to create genome-wide reduction-of-function modifications requires only a single step of transformation of the host cells[19, 20] (**Fig. 3.1B**). Such simplicity and

effectiveness should enable the use of directed evolution strategy by repeating the cycles of RNAi screening to accumulate beneficial knockdown modifications (**Fig. 3.1C**). RNAi is a cellular gene silencing mechanism broadly distributed in eukaryotic organisms, whereby messenger RNAs (mRNAs) are targeted for degradation by homologous double-stranded RNAs (dsRNAs)[21, 22]. RNAi screening enables genome-wide reduction-of-function perturbations without allelic modifications, and is widely used in eukaryotic functional genomics[20, 23]. All known *S. cerevisiae* strains lack a native RNAi pathway[24]. Recently, a heterologous RNAi pathway from *Saccharomyces castelli* was functionally reconstituted in *S. cerevisiae* to achieve effective gene silencing[24]. Three human proteins, Ago2, Dicer and TRBP, were also found to be sufficient to enable gene knockdown in *S. cerevisiae* by RNAi[25]. In addition, the RNAi machinery was implemented as a metabolic engineering tool to improve itaconic acid production in this yeast[26]. So far, however, no RNAi screening has been reported in this model eukaryotic organism. We first demonstrated RNAi screening in *S. cerevisiae* for suppressor analysis of a telomerase-deficient mutation *yku70*Δ, and then applied iterative RNAi screening for improved acetic acid (HAc) tolerance.

3.2. Results

3.2.1. Functional reconstitution of the RNAi pathway

We first sought to establish RNAi screening in *S. cerevisiae*. A *Saccharomyces castellii* RNAi pathway was functionally introduced into *S. cerevisiae* recently[24]. We reconstituted this *S. castellii* RNAi pathway into our target yeast strain CEN.PK2-1c and named the resulting strain as the CAD strain (**Fig. 3.2**).

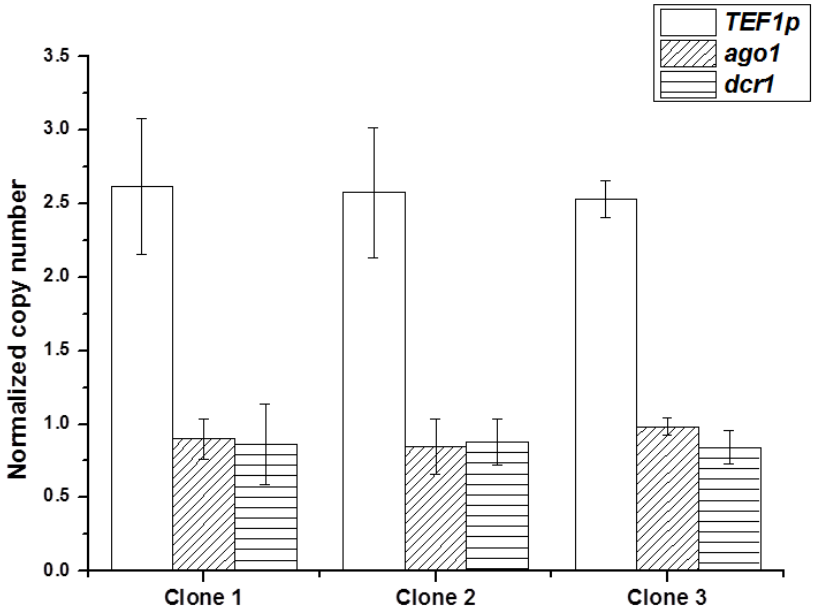

Figure 3.2. The integration copy number of the RNAi pathway in the CAD strain determined by quantitative real-time PCR (qPCR). Three colonies were assayed. The genomic DNA of the CAD strain was used as template, and the primers qTEF1p For/Rev, qAgo1 For/Rev and qDcr1 For/Rev were used to quantify the copy numbers of P_{TEF1}, *ago1* and *dcr1*. The *alg9* gene was used as the internal control whose copy number was set to be one. Error bars represent the two concentrations of genomic DNA (differed by 10 fold) added as template in the qPCR reactions. The integration copy number of the RNAi pathway was determined to be one (for P_{TEF1}, there is an endogenous copy in the genome, so the total copy number will be two).

We observed repression of the expression of a reporter protein, green fluorescent protein (GFP), to various degrees by the *gfp*-silencing constructs based on either the anti-sense design or the convergent-promoter design (**Fig. 3.3**). The TEF1p-GFPrc construct transcribes the full-length anti-sense RNA of the GFP gene, while the TTrc-GFP constructs transcribe dsRNAs derived from different regions of the GFP gene (**Fig. 3.3A**). The reduction of GFP fluorescence was dependent on both the RNAi pathway and the RNA expression (**Fig 3.3**), indicating that the imported RNAi pathway was functional. For the convergent-promoter constructs, both the position and the length of the inserts affected the repression efficiency. In particular, the 223-631 bp region corresponds to the digestion product of the GFP gene by *Sau*3AI, and the derivative dsRNA led to a 40% reduction of GFP fluorescence signal. DNA fragments derived from 1-180 bp, 1-360 bp, 1-540 bp and 1-717 bp of the GFP gene were also cloned into the convergent promoter cassette, and it was found that the repression efficiency decreased with longer insets (**Fig. 3.3B**).

The convergent-promoter design was subsequently used to construct a genome-wide RNAi library. Convergent promoters can drive the *in vivo* synthesis of two complementary RNA molecules to form a dsRNA molecule. The dsRNA can trigger the RNaseIII-like enzyme Dicer to cleave long dsRNAs into small interferencing RNAs (siRNAs) of 21-25 bp. These siRNAs are then loaded into a multi-subunit RNA-induced silencing complex (RISC), which will target and degrade the cellular mRNA with homologous sequences. In particular, two strong constitutive promoters, P_{TEF1} and P_{TPI1}, were used as the convergent pair, between which a *Bam*HI site was engineered to facilitate the insertion of genomic DNA fragments generated by *Sau3*AI digestion. The convergent promoter cassette was assembled into a single-copy plasmid

pRS416 to construct the plasmid pRS416-GPDtrc-TEF1p-*Bam*HI-TPI1prc- PGK1t (aka pRS416-TTrc).

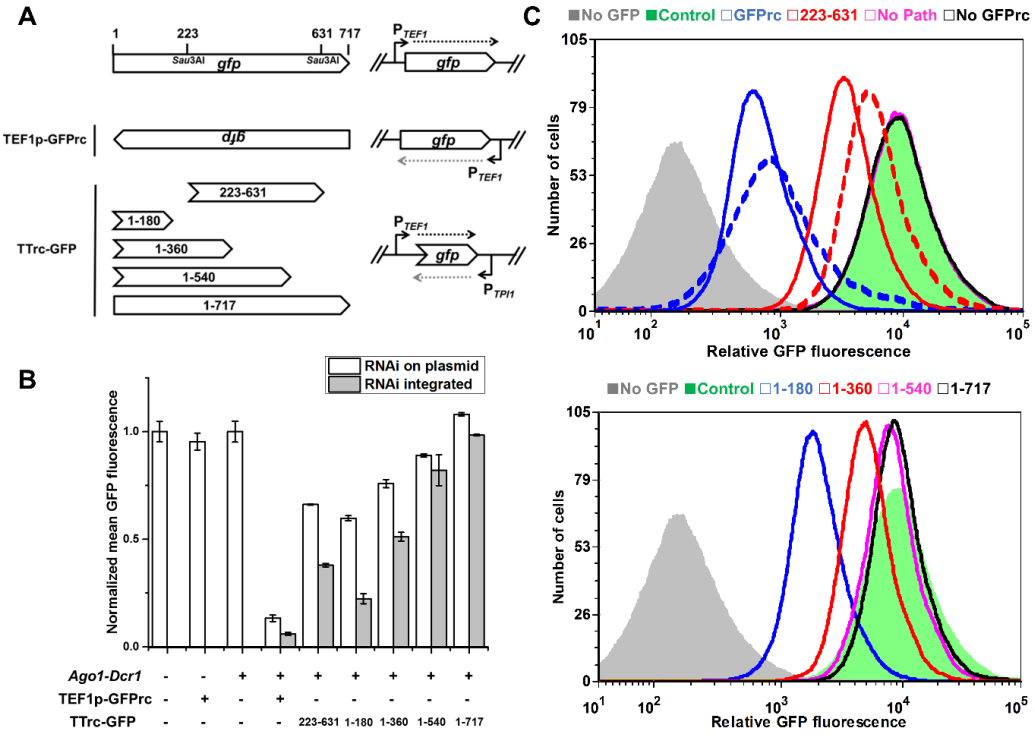

Figure 3.3. Repression of GFP fluorescence by RNAi constructs. The overexpression cassette of GFP was integrated into the *leu2* locus of the CAD or CEN.PK2-1c strain. (**A**) Scheme for RNAi expression design. (**B**) Silencing of GFP expression. The average on the mean GFP fluorescence of three biological replicates are reported. The error bars indicate standard deviations. The 100% reference of GFP signal was defined as the strain containing the integrated GFP expression cassette and the control plasmid pRS416. The RNAi cassettes locate either on a single-copy plasmid pRS416 (white bar) or in the *ura3* locus of the yeast genome (grey bar). (**C**) FACS histograms showing GFP fluorescence of the yeast strains harboring different RNAi constructs. Dashed lines indicate the plasmid pRS416 is used to express the RNAi cassettes, whereas solid lines indicate the RNAi cassettes are integrated in the *ura3* locus. The *No GFP* strain is the CEN.PK2-1c strain with the pRS416-TTrc plasmid. The *Control* strain is the CEN.PK2-1c strain with an integrated GFP-overexpressing cassette and the pRS416-TTrc plasmid. The *No Path* strain is the CEN.PK2-1c strain with an integrated GFP-overexpressing cassette and the pRS416-TEF1p-GFPrc plasmid. The *No GFPrc* strain is the CAD strain with an integrated GFP-overexpressing cassette and the empty pRS416 plasmid.

By adapting the convergent-promoter design (**Fig. 3.3A**), we created a pooled long-dsRNA library from the yeast genomic DNA fragments which were generated by enzymatic digestion[27]. We randomly sequenced 50 plasmids from the library, and the statistical analysis of the DNA sequencing results was summarized in **Fig. 3.4**. The insets of the majority of plasmids consisted of only one genomic fragment, while for some contained more than one DNA fragments (**Fig. 3.4B**), probably resulting from the ligation of genomic DNA fragments with compatible ends. To prevent the self-ligation between the genomic DNA fragments, a fill-in reaction of the overhangs could be performed by Klenow fragment. Together, 83 fragments were present in the 50 sequenced plasmids. The sizes of the fragments ranged from 41 bp to 1109 bp, with an average size of 177 bp (**Fig. 3.4A**).

Table 3.1. Estimation of library coverage

Library Size	Redundancy	Probability of coverage*
50,000**	0.71	50.76%
100,000**	1.42	75.75%
422,400[#]	5.98	99.75%
500,000**	7.08	99.92%

***** The coverage was calculated based on the assumption of the Poisson distribution. The equation was $E < C >= 1\text{-}\exp(L \times N/G)$[28], where $E<C>$ is the probability of any single nucleotide in the genome covered by at least one clone in the library, L is the average length of the fragment which is equal to 177 bp, N is the library size, and G is the genome size of *S. cerevisiae* which is 12.5 Mbp.

** Simulated calculation by assuming the library size equals to 50, 000, 100, 000 and 500, 000

[#] The actual library size obtained for the pooled RNAi library on a plasmid

This 177 bp average fragment size was employed to estimate the coverage of the yeast genome by the genomic DNA derived RNAi library (**Table 3.1**). The possibility of covering the

complete yeast genome was estimated at more than 99.8% (**Table 3.1**). The sources of these fragments were randomly distributed on the yeast genome, including 17 out of 18 yeast chromosomes, the mitochondrial DNA and the 2μ plasmid (**Fig. 3.4C**). Taken together, it was suggested that a satisfying coverage of the yeast genome was achieved.

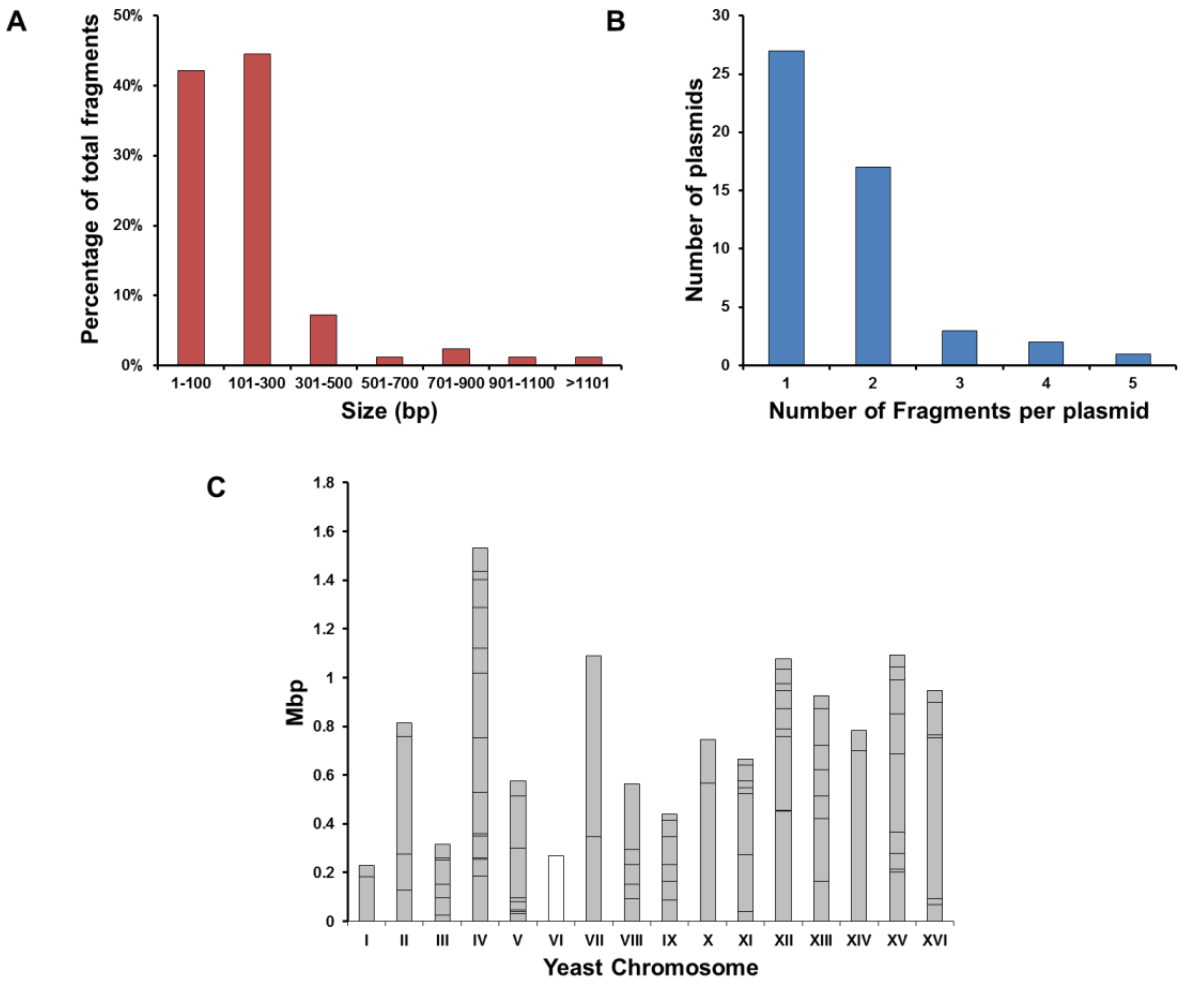

Figure 3.4. The DNA sequencing results of 50 randomly picked plasmids from the RNAi library. (**A**) The size distribution of the genomic DNA fragments. (**B**) The number of fragments per plasmid. (**C**) The locations of the genomic DNA mapped to the yeast genome. Each column represents one chromosome, the height of which is proportional to the sizes of the chromosomes. Each bar indicates the location of one fragment. The fragments derived from the mitochondrial genome and 2μ plasmid were omitted for clarity.

3.2.2. Suppressor analysis of the yku70Δ mutation by RNAi screening

To verify the RNAi library, we performed a genome-wide RNAi screening to identify suppressors of the *yku70*Δ mutation (**Fig. 3.5A**). Yku70 is a telomere-associated protein[29, 30]. At an elevated temperature such as 37 °C, the null mutation of the *yku70* gene will lead to single stranded DNA accumulation at the telomere, which triggers DNA damage response and then cell-cycle arrest[29]. The temperature-dependent growth phenotype caused by the *yku70*Δ mutation provides a valuable model to study genetic interactions involved in DNA-repair and cell-cycle pathways[29-31]. We transformed our RNAi library plasmids into the CAD strain with the *yku70* gene deleted. The transformants were selected on solid synthetic medium at 37 °C. Colonies that grew larger than the control strain were picked up and tested for their growth capacity under challenging temperatures in liquid media. Plasmids isolated from the top growers were re-transformed into fresh yeast cells, and those still conferring improved growth at 37 °C were sequenced to identify the origins of the RNAi cassettes (**Fig. 3.5B**). The inserts within the selected RNAi cassettes were found to be the fragments from one essential gene *ret1*, and four non-essential genes *say1*, *ssa1*, *cst6* and *mlp2* (**Table S2**). The *mlp2* and *ssa1* genes were previously identified as suppressors of the *yku70*Δ mutation[31, 32], which confirmed the effectiveness of our RNAi library for genome-wide screening.

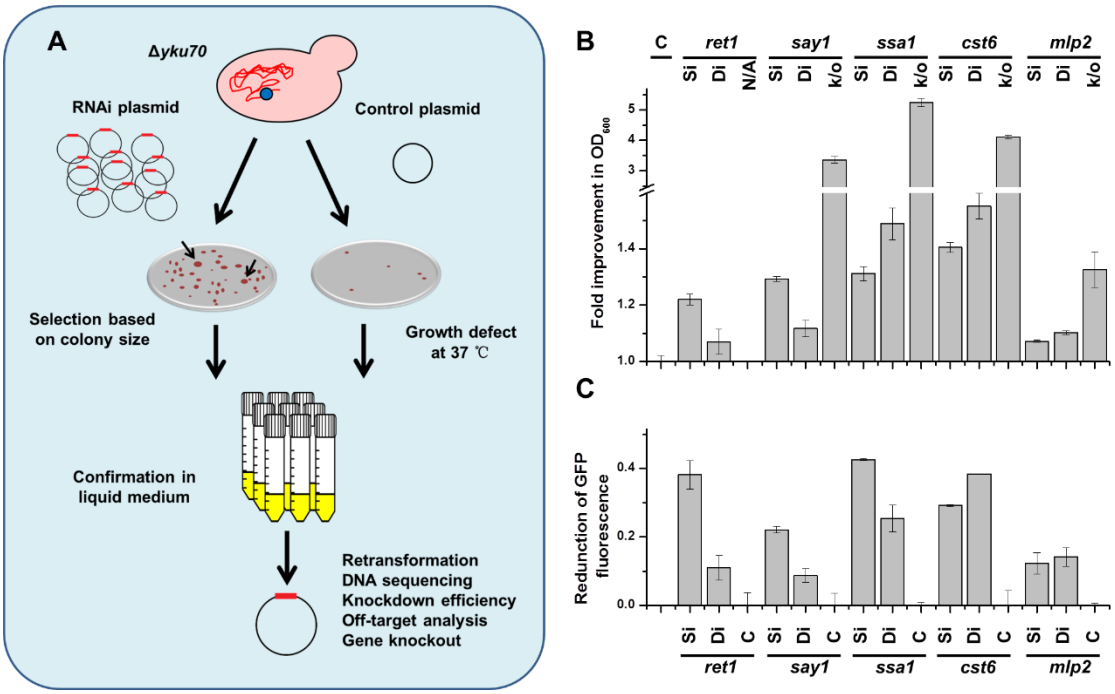

Figure 3.5. RNAi screening in *S. cerevisiae* to identify *yku70*Δ suppressors. (**A**) Scheme for suppressor screening of *yku70*Δ. (**B**) Comparison of growth capacity in synthetic dropout medium at 37 °C. The initial OD_{600} was 0.2, and the cell density after growing for 12 hrs was normalized to the CAD strain containing the control plasmid. (**C**) Estimation of knockdown efficiency where the expression levels of target proteins were quantified by the GFP tag[33]. The 100% reference of GFP signal was defined as the strains with the control plasmid for each target gene respectively. Reduction of GFP fluorescence was reported as 1-(RNAi strain fluorescence / Control strain fluorescence). All the RNAi constructs were transcribed from a single-copy plasmid pRS416. Si: selected RNAi constructs; Di: designed RNAi constructs; k/o: knockout; C: control plasmid pRS416-TTrc; N/A: knockout is lethal. Error bars indicate standard deviation of three biological replicates.

To eliminate the "off-target" effects, we designed a second RNAi construct targeting a different region of the same transcript for each gene (**Table 3.2**). The "off-target" effect often resulted from the partial homology to other transcripts[34], and it was unlikely that two independent RNAi cassettes would have the common "off-target" suppression effects. All the designed RNAi constructs rescued the temperature sensitivity of *yku70*Δ at 37 °C (**Fig. 3.5B**).

We further estimated the knockdown efficiency by fusing the target endogenous proteins with a carboxyl-terminal GFP tag for quantification[33]. All the selected and designed RNAi cassettes showed reduction of GFP fluorescence (**Fig. 3.5C**). The suppression effects of the selected non-essential genes were also validated by examining the gene knockout mutants (**Fig. 3.5B**). These results confirmed that the selected genes were suppressors of the *yku70*Δ mutation.

Table 3.2. Summary of the selected and designed RNAi cassettes for HAc tolerance

Name	Number of fragments per inserts	Size (bp)	Mapped to the target gene
*ret1*_Si	1	75	*ret1* 2803-2877/3450*
*say1*_Si	1	123	*say1* 91-213/1275
*ssa1*_Si	1	67	*ssa1* 213-279/1929
*cst6*_Si	2**	46	*cst6* 1629-1584/1764
		70	*ydr109c 100-169/*2148
mlp2_Si	2	120	*mlp2* 2937-2818/5040
		51	V 552895-552945/1091291***
*ret1*_Di	1	180	*ret1* 662-841/3450
*say1*_Di	1	180	*say1* 801-980/1275
*ssa1*_Di	1	180	*ssa1* 488-667/1929
*cst6*_Di	1	180	*cst6* 311-490/1764
mlp2_Di	1	180	*mlp2* 101-280/5040

* For the genes, A-B/X indicated that the RNAi insert corresponds to the A-B bp region of the gene's ORF, whose length is X bp.

** For some of the selected RNAi plasmids, the inserts containing more than one inserts. In this case, a designed RNAi construct as well as the knockout mutant were created to confirm which RNAi cassette(s) within the inserts was (were) the positive hit(s).

*** If the fragment is not from an ORF, the location is denoted as a reference to the chromosome.

While the newly-isolated modifiers will provide novel molecular insights on telomere biology (**Table 3.3**), the identification of an essential gene as a suppressor of the *yku70*Δ

mutation highlighted the advantage of RNAi screening over gene-knockout libraries by including the genes whose null mutations are lethal.

Table 3.3. Gene functions of confirmed suppressor of *yku70*Δ*.

Gene	Description (www.yeastgenome.com)
ret1	Second-largest subunit of RNA polymerase III, which is responsible for the transcription of tRNA and 5S RNA genes, and other low molecular weight RNAs
say1	Sterol deacetylase; component of the sterol acetylation/deacetylation cycle along with Atf2p; integral membrane protein with active site in the ER lumen; green fluorescent protein (GFP)-fusion protein localizes to the endoplasmic reticulum
ssa1	ATPase involved in protein folding and NLS-directed nuclear transport; member of HSP70 family; forms chaperone complex with Ydj1p; localized to nucleus, cytoplasm, and cell wall; 98% identical with paralog Ssa2p, but subtle differences between the two proteins provide functional specificity with respect to propagation of yeast [URE3] prions and vacuolar-mediated degradations of gluconeogenesis enzymes; general targeting factor of Hsp104p to prion fibrils
cst6	Basic leucine zipper (bZIP) transcription factor, in ATF/CREB family; mediates transcriptional activation of NCE103 (encoding carbonic anhydrase) in response to low CO2 levels such as in the ambient air; proposed to be a regulator of oleate responsive genes; involved in utilization of non-optimal carbon sources and chromosome stability; CST6 has a paralog, ACA1, that arose from the whole genome duplication
mlp2	Guanine nucleotide exchange factor (GEF) for ADP ribosylation factors involved in proliferation of the Golgi, intra-Golgi transport and ER-to-Golgi transport; found in the cytoplasm and on Golgi-associated coated vesicles

***** It is noteworthy that the *ssa1* and *mlp2* genes were known suppressors of *yku70*Δ mutation. The deletion of *cst6* was reported to rescue the temperature sensitivity of the *cdc13-1* mutant strain[35], where the *cdc13* gene encodes another telomere-capping protein[29]. The Say1 protein physically interacts with Sap30[36], which was also identified as a suppressor of *yku70*Δ[31], suggesting *say1* and *sap30* were involved in the same regulatory mechanism in response to *yku70*Δ defect.

3.2.3. RNAi-assisted genome evolution for improved HAc tolerance

After establishing RNAi screening in *S. cerevisiae*, we combined it with directed evolution to rapidly engineer yeast cells for improved HAc tolerance (**Fig. 3.1C**). Although bioethanol fermentation by *S. cerevisiae* from sugarcane sucrose and corn starch has been widely applied for biofuel production[9, 10], the use of lignocellulosic biomass as substrate by this yeast is highly desirable for a more sustainable biofuel process[37]. Acetic acid (HAc) is an unavoidable inhibitor either from the pretreatment step of lignocelluloses or as the by-product during alcoholic fermentation[38, 39], and tolerance to HAc is highly desirable for commercial production of chemicals and fuels by *S. cerevisiae*[39].

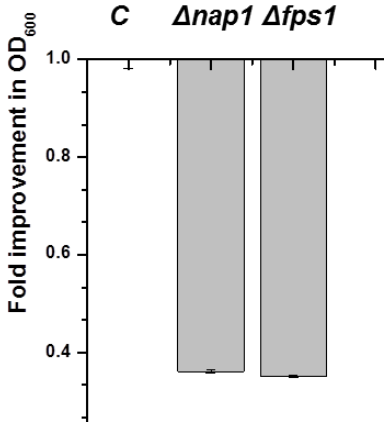

Figure 3.6. Reduced HAc tolerance of mutant strains *nap1*Δ and *fps1*Δ in CEN.PK2-1c background. The strains were cultured aerobically in SC-U medium containing 0.5% (v/v) HAc. The initial OD_{600} was 0.2, and the cell density after growing for 12 hrs was normalized to the CAD strain containing the control plasmid. Error bars indicate standard deviation of three biological replicates. The introduction of *nap1*Δ or *fps1*Δ was known to improve HAc tolerance in the BY4741 strain[40, 41], but failed to elicit the same phenotype in the CEN.PK2-1c strain.

Previous screening efforts to increase HAc tolerance in *S. cerevisiae* are limited to BY4741/4742 strains based on which most gene-knockout libraries are created[39, 42]. Our results

showed that the mutations that were reported to improve HAc tolerance in BY-strains failed to elicit growth advantage in our CEN.PK2-1c strain under HAc stress (**Fig. 3.6**).

We first confirmed that introduction of this heterologous RNAi pathway had little impact on cellular growth and HAc tolerance of *S. cerevisiae* (**Fig. 3.7A**), which is consistent with a previous report[43].

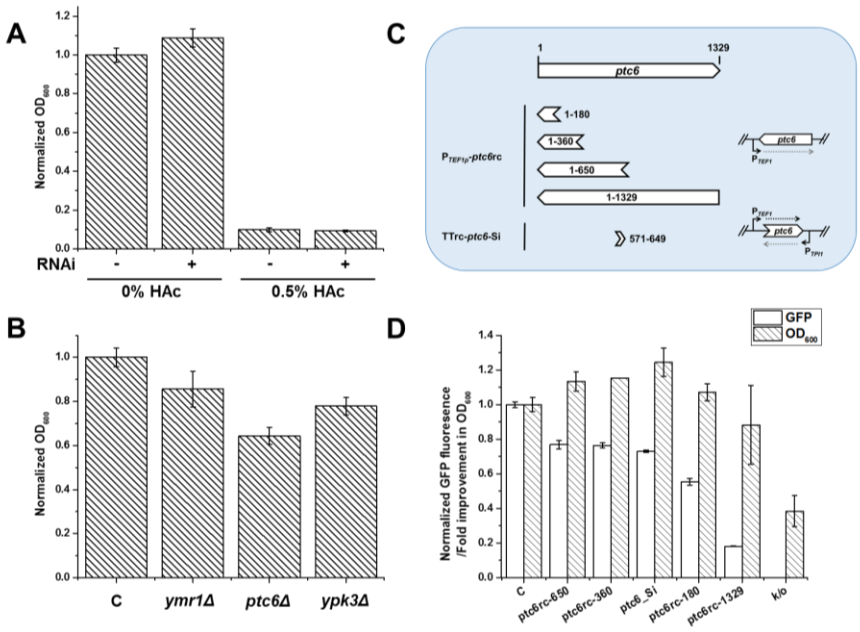

Figure 3.7. Improving HAc tolerance by RNAi screening in *S. cerevisiae*. (**A**) No obvious difference in growth capacity was found upon introduction of the *S. castellii* RNAi pathway. The strains were grown aerobically in synthetic dropout medium (pH=4.5) with or without 0.5% (v/v) HAc. The initial OD_{600} was 0.2, and the cell density after growing for 12 hr was normalized to the wild-type strain containing the control plasmid. The same culture condition was employed in (**B**) and (**D**). (**B**) The knockout mutations of the selected gene targets in the first round of RAGE led to reduced growth rates in synthetic dropout medium without HAc stress compared to the CAD strain containing the control plasmid. (**C**) Scheme for RNAi expression cassettes of the *ptc6* gene. The *ptc6* gene fragments with varying lengths were cloned in the anti-sense direction under the control of P_{TEF1} promoter. TTrc-*ptc6*-Si was the selected RNAi cassette with the convergent-promoter design. (**D**) The different knockdown levels enabled by various RNAi cassettes of the *ptc6* gene led to varying HAc tolerance in the recombinant yeast strains. Error bars indicate standard deviation of three biological replicates. The knockdown efficiency was estimated by the GFP-tagging experiment following the same procedure as in **Fig. 3.8B.**

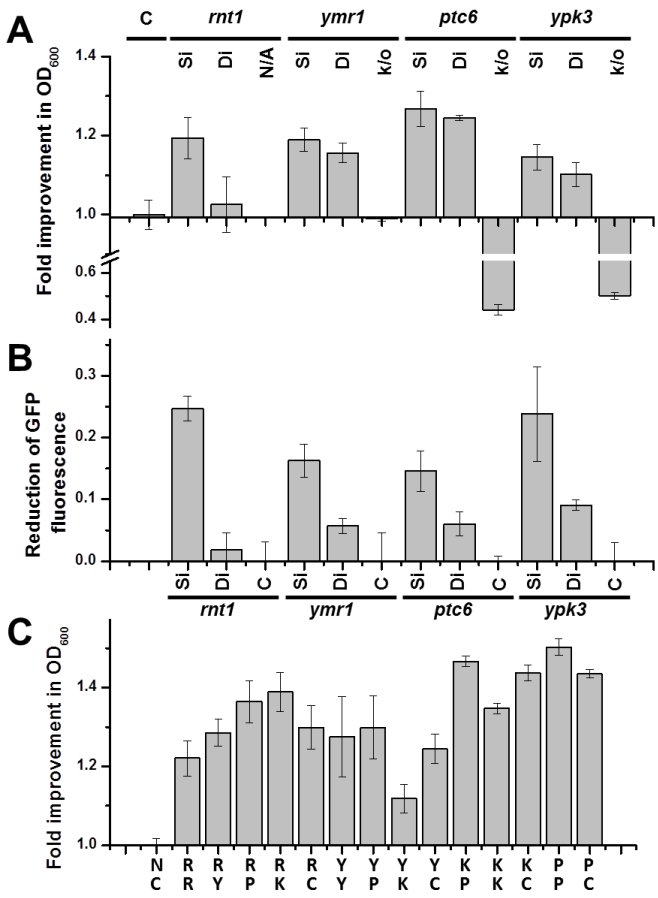

Figure 3.8. Knockdown targets identified by RNAi screening for improved HAc tolerance. (**A**) Comparison of growth capacity in the presence of 0.5% (v/v) HAc in synthetic dropout medium (pH = 4.5). The initial OD_{600} was 0.2, and the cell density after growing for 12 hrs was normalized to the CAD strain containing the control plasmid. (**B**) Estimation of knockdown efficiency where the expression levels of target proteins were quantified by the GFP tag[33]. The 100% reference of GFP signal was defined as the strains with the control plasmid for each target gene respectively. Reduction of GFP fluorescence was reported as 1-(RNAi strain fluorescence/ Control strain fluorescence). All the RNAi constructs were transcribed from a single-copy plasmid pRS416 in (**A**) and (**B**). (**C**) HAc tolerance of strains with combinations of RNAi cassettes from the first round of RAGE. The same growth condition was employed as in (**A**). The names of the strains are denoted by two letters. The first letter indicates the integrated cassette in the *his3* locus, while the second letter indicates the RNAi cassette on the plasmid. R: *rnt1*_Si, Y: *ymr1*_Si, P: *ptc6*_Si, K: *ypk3*_Si, N: no integration C: the control plasmid pRS416-TTrc.

Table 3.4. Summary of the selected and designed RNAi cassettes for HAc tolerance

Name	Number of fragments per inserts	Size (bp)	Mapped to the target gene
Based on CAD			
A12 (*rnt1*_Si)	1	152	*rnt1* 991-840/1416*
D11 (*ymr1*_Si)	1	119	*ymr1* 265-147/2067
F6 (*ptc6*_Si)	1	79	*ptc6* 649-571/1329
H6 (*ypk3*_Si)	1	1017	*ypk3* -79-937/1578
*rnt1*_Di	1	180	*rnt1* 457-636/1416
*ymr1*_Di	1	180	*ymr1* 1631-1810/2067
*ptc6*_Di	1	180	*ptc6* 721-900/1329
*ypk3*_Di	1	180	*ypk3* 471-650/1578
Based on CAD*_ymr1i*			
D11_#9	1	142	*sec7* 3677-3536/6030
Based on CAD*_ptc6i*			
F6_#19	1	198	*ypr084w* 365-168/1371
F6_#23	1	103	*afg3* 2286-2236/2286
Based on CAD*_ptc6i_ypr084wi*			
F6_#19_#18	1	61	*rgd1* 1342-1282/2001
F6_#19_#27	1	72	XIII 586652-586580/ 924431**, including part of the $tRNA^{Val(\underline{A}AC)}$

* For the genes, A-B/X indicated that the RNAi insert corresponds to the A-B bp region of the gene's ORF, whose length is X bp.

** If the fragment is not from an ORF, the location is denoted as a reference to the chromosome.

Following a similar procedure of the suppressor screening for *yku70*Δ (**Fig. 3.6A**), four RNAi cassettes that conferred yeast cells with increased growth under 0.5% (v/v) HAc stress were identified and confirmed in the first round of RAGE (**Fig. 3.8A** and **Fig. 3.8B**, and **Table 3.4**). We then integrated the four selected RNAi cassettes into the yeast genome separately, and the resultant strains all exhibited better fitness than the wild-type strain under HAc stress (**Fig. 3.8C**). These strains were then employed as new parent strains for the second round of RNAi screening (**Fig. 3.9A**). With a mini-library containing the four selected plasmids from the first round, we found that the combination of the beneficial RNAi modifications did not necessarily lead to

incremental improvements (**Fig. 3.8C**), highlighting the necessity to perform genome-wide screening to identify targets for further engineering.

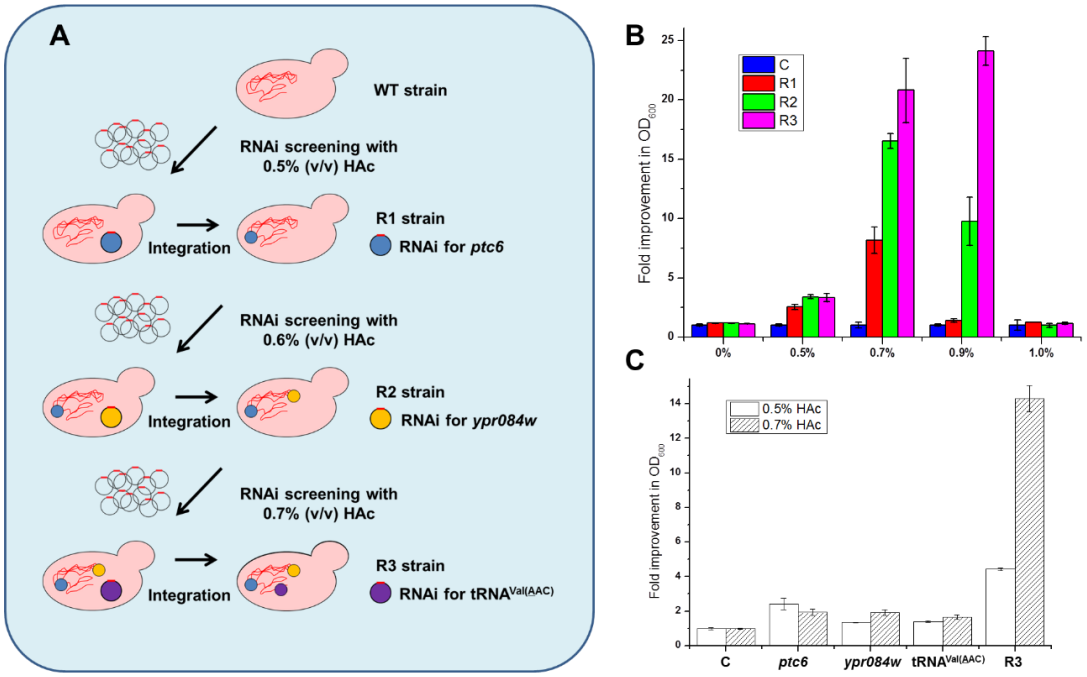

Fig. 3.9. Yeast strains engineered by RAGE showing improved HAc tolerance. (**A**) Scheme of iterative RNAi screening to accumulate beneficial knockdown modifications in a yeast genome. (**B**) Growth capacity of the yeast strains identified from three rounds of RAGE with different levels of HAc in synthetic dropout medium (pH = 4.5). The initial OD_{600} was 0.01, and the cell density was measured after growing for 24 hrs (0%, 0.5% and 0.7% (v/v)) or 48 hrs (0.9% and 1.0%). No cell growth was observed for the 1.0% (v/v) HAc group after 72 hrs. Fold improvements were compared to the CAD strain containing the control plasmid. (**C**) Contribution of individual knockdown modification from the R3 strain to the enhanced HAc tolerance. Plasmids harboring the selected RNAi cassettes for *ptc6*, *ypr084w* and $tRNA^{Val(\underline{A}AC)}$ were transformed into the CAD strain to achieve individual gene silencing. The same growth condition was employed as in (**C**) except that only two HAc concentrations (0.5% and 0.7%) were tested. C: the CAD strain harboring the control plasmid pRS416-TTrc; R1-R3: the best strains selected from the first, second and third round of RAGE, respectively, with all the selected RNAi cassettes integrated (see **Table 3.8** for details). The R1-R3 strains were all transformed with the control plasmid pRS416-TTrc to enable growth in the SC-U medium. Error bars indicate standard deviation of three biological replicates.

The genome-wide RNAi library was transformed into the new parent strains (**Fig. 3.9A**). A higher HAc concentration (0.6% (v/v)) than the first round (0.5% (v/v)) was used to isolate mutant strains with better HAc tolerance. In the background of the strain integrated with the RNAi cassette for *ptc6* (named as the first round strain R1), the RNAi cassette for *ypr084w* resulted in the highest fitness in the second round of screening (**Fig. 3.9A**). Thus the recombinant strain integrated with both the *ptc6* RNAi and *ypr084w* RNAi cassettes (named as the second round strain R2) was created and used as the parent strain for the third round of RNAi screening with 0.7% (v/v) HAc as selection pressure. The best mutant strain was found to contain an RNAi cassette with a genomic DNA fragment from $tRNA^{Val(\underline{A}AC)}$ as the insert (named as the third round strain R3, **Fig. 3.9A**).

3.2.4. Superior HAc tolerance of engineered strains by RAGE

The mutant strains with the highest fitness in each round of RAGE were compared in parallel with the control strain for the growth capacity in the presence of different concentrations of HAc. A general trend of R3 > R2 > R1 > control was observed under almost all the conditions, indicating the step-wise improvement of HAc tolerance (**Fig. 3.9B**). The tolerance phenotype of the R3 strain was significantly increased, which accumulated greater than 20-fold more biomass under 0.9% (v/v) HAc relative to the control strain. The 100% inhibitory HAc concentration for growth was elevated from 0.8% to 1.0% (v/v).

Next, we evaluated the contribution of each RNAi cassette to the HAc tolerance of the R3 strain. The yeast strains carrying only one of the selected RNAi cassettes for *ptc6*, *ypr084w* and $tRNA^{Val(\underline{A}AC)}$ were constructed and compared with the R3 strain for the growth capacity with elevated HAc levels (0.5% and 0.7% (v/v)). The results showed that each cassette led to

increased biomass accumulation to some extent relative to the control, but the superior tolerance of the R3 strain cannot be readily explained by simply adding up these individual effects (**Fig. 3.9C**). The mechanism of the observed synergy cannot be deduced directly from the known functions of the targeted genes (**Table 3.5**). Such non-linear interactions between the RNAi cassettes highlighted the necessity of iterative rounds of screening to accumulate the beneficial knockdown perturbations.

Table 3.5. Gene functions of the knockdown targets identified by RAGE.

Gene	Description (www.yeastgenome.com)
rnt1	Nuclear dsRNA-specific ribonuclease (RNase III); involved in rDNA transcription, rRNA processing and U2 snRNA 3' end formation by cleavage of a stem-loop structure at the 3' end of U2 snRNA; involved in polyadenylation-independent transcription termination; involved in the cell wall stress response, regulating the degradation of cell wall integrity and morphogenesis checkpoint genes
ymr1	Phosphatidylinositol 3-phosphate (PI3P) phosphatase; involved in various protein sorting pathways, including CVT targeting and endosome to vacuole transport; has similarity to the conserved myotubularin dual specificity phosphatase family
ptc6	Mitochondrial type 2C protein phosphatase (PP2C) with similarity to mammalian PP1Ks; involved in mitophagy; null mutant is sensitive to rapamycin and has decreased phosphorylation of the *Pda1* subunit of pyruvate dehydrogenase
ypk3	An AGC kinase phosphorylated by cAMP-dependent protein kinase (PKA) in a TORC1-dependent manner
sec7	Guanine nucleotide exchange factor (GEF) for ADP ribosylation factors involved in proliferation of the Golgi, intra-Golgi transport and ER-to-Golgi transport; found in the cytoplasm and on Golgi-associated coated vesicles
ypr084w	Putative protein of unknown function
afg3	Component, with *Yta12p*, of the mitochondrial inner membrane m-AAA protease; mediates degradation of misfolded or unassembled proteins and is also required for correct assembly of mitochondrial enzyme complexes; involved in cytoplasmic mRNA translation and aging
rgd1	GTPase-activating protein (RhoGAP) for *Rho3p* and *Rho4p*, possibly involved in control of actin cytoskeleton organization
$tRNA^{Val(\underline{A}AC)}$	tV(AAC)M3, Valine tRNA (tRNA-Val), predicted by tRNAscan-SE analysis

The fermentation performance of the R3 strain and the control strain were evaluated under an oxygen-limited condition. Three levels of HAc stress were applied as 0%, 0.7% and 1.0% (v/v) in the synthetic minimal medium. The performance of the R3 strain was superior to the control under all the conditions in terms of glucose utilization, biomass accumulation and ethanol productivity (**Fig. 3.10**). Specifically, in the presence of 0.7% HAc, the R3 strain exhibited a 63.1±3.2% increase in ethanol productivity relative to the control strain in the period of 0-24 hr. Though the final ethanol titers were similar for the two strains, the R3 strain accumulated 15.3±3.5% more biomass than the control. Without the HAc stress, the R3 strain showed the same ethanol production profile as the control strain. Notably, after the glucose was depleted, the R3 strain exhibited a much faster assimilation rate of acetate than the control (**Fig. 3.10**). The enhanced capacity to utilize acetate might provide some clues why the R3 strain had a higher level of HAc tolerance. These results confirmed that RAGE successfully improved the HAc tolerance and fermentation performance of *S. cerevisiae*.

3.3. Conclusions and Future Prospects

RNAi proves to be an enabling technology with broad applications for functional analysis, therapeutics and metabolic engineering in mammalian cells, insect cells and plants[44, 45]. Although RNAi is not a novel technique, RAGE further expands the power of RNAi by the identification and fine-tuning of multiplex gene targets and engineer yeast cells on a genome scale. The potential of employing RNAi screening for directed genome evolution offers great advantages over conventional gene-knockout strategies, such as providing a simple tool to modify a eukaryotic genome globally and iteratively, enabling fine-tuning of gene expression, including essential genes in functional screening, and identifying beneficial traits which requires

synergistic genetic modifications. By targeting mRNAs to introduce reduction-of-function mutations, RAGE might be especially useful in industrial organisms for which gene deletion is extremely challenging due to polyploidy[46].

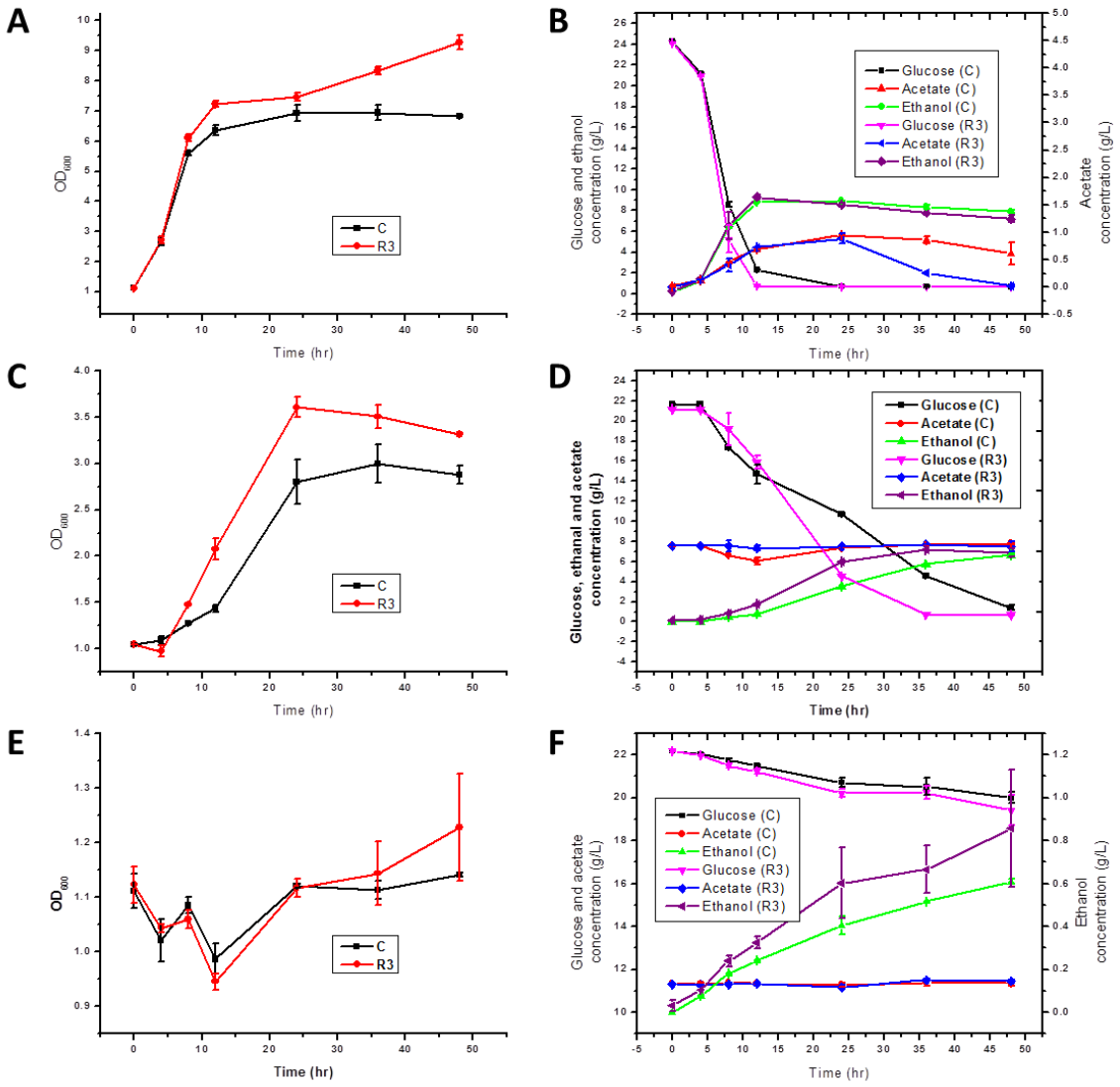

Figure 3.10. Fermentation profiles of the engineered and wild-type yeast strains. Cells were cultured in biological replicates in the SC-U medium containing 20 g/L glucose and 0% (**A** and **B**), 0.7% (**C** and **D**) and 1% (**E** and **F**) HAc (v/v). The initial OD_{600} was adjusted to 1. Oxygen-limited condition was employed with un-baffled shake flasks and 100 rpm agitation. Biomass accumulation was monitored by optical density (**A**, **C** and **E**). The concentrations of glucose, acetate and ethanol were measured by HPLC (**B**, **D** and **F**). Error bars indicate standard deviation of three biological replicates.

When interpreting the results of RNAi screening, cautions need to be taken against "off-target" effect, which is mainly resulted from cross-silencing of other transcripts with partial homology[34]. The "off-target" effect is well-documented for different eukaryotic systems in literature[20, 34, 47, 48]. The most effective way to ensure that the observed phenotypes are "on-target" is to show that these phenotypes can be generated by independent RNAi molecules which target the same gene but contain completely distinct sequences[19, 20, 23], as performed in this study by analyzing the phenotypes caused by a second designed dsRNA molecule for each gene target (**Fig. 3.5B** and **Fig. 3.8A**). While it is very critical to minimize "off-target" effect when applying RNAi as therapeutics[49], it is probably less important to achieve high specificity of RNAi in microbial strain engineering.

In this work, we constructed the RNAi library by inserting the genomic DNA fragments generated by *Sau*3AI into a pair of convergent promoters for dsRNA synthesis. Whereas this is an established method[27], it is possible to further improve the coverage and effectiveness of the RNAi library. First, though a library size that is big enough can ensure the complete coverage of the yeast genome, the biased fragmentation pattern determined by the recognition site of *Sau*3AI may preclude the identification of some genes in the screening. For these genes, the dsRNA molecules transcribed from the *Sau*3AI fragments may not generate observable phenotypes due to weak knockdown levels. To solve this issue, a collection of random fragmentation strategies with less inherent bias through enzymatic, chemical and mechanical means can be employed, which are originally developed to create the shotgun libraries for genome sequencing[50]. Second, the knockdown efficiency enabled by the long dsRNA molecules is moderate, which may limit the sensitivity of RNAi screening. For example, the knockout mutations of some identified *yku70*Δ suppressors led to more substantial growth advantage than the knockdown mutations

(**Fig. 3.3B**). Whereas the incompleteness of knockdown mutations is a well-recognized issue of RNAi screening, it is possible to improve the gene-silencing efficiency by optimizing the format of RNAi reagents. For example, in this study we observed that the full-length anti-sense RNA resulted in more profound knockdown effects than the dsRNA molecules transcribed by the convergent promoters (**Fig. 3.3B**). Therefore, one way to improve the sensitivity of RNAi screening is to use the full-length anti-sense RNAi library, which can be created by cloning of the full-length cDNA library in a reversed direction after a promoter sequence (**Chapter 4**). It is noted that in a recent report, hairpin RNAi molecules were implemented for RNAi down-regulation in *S. cerevisiae*[26]. Different parameters, such as the hairpin length and expression context, were manipulated for optimized gene silencing effectiveness. It is thus possible to use hairpin RNAi design for the library construction in RAGE. For *S. cerevisiae* strains with sequenced genomes, it is also possible to design synthetic genome-wide RNAi library by computational algorithms to optimize the potency and specificity[51].

In its current form, a pooled library on an episomal plasmid coupled with an efficient selection strategy is implemented in RAGE. However, a separate integration step is needed at the beginning of each round to create a new parent strain. It should be noted that the plasmid-borne and integrated versions of the same RNAi cassette may lead to different phenotypes. For example, the integrated RNAi cassette for *ptc6* (**Fig. 3.8C**, group PC) resulted in a slightly higher HAc tolerance than the plasmid-borne version (**Fig. 3.8A**, group *ptc6*_Si). In addition, the *gfp*-knockdown constructs that were integrated generally led to a stronger silencing effect than their plasmid-borne counterparts (**Fig. 3.3B**). The strains harboring the integrated RNAi cassettes also exhibited less population heterogeneity in GFP fluorescence profile, compared to the strains with the RNAi cassettes on the plasmid (**Fig. 3.3C**). Together, these observations

indicate that the different expression contexts, from a plasmid or a genomic location, may affect the kinetics of dsRNA transcription and therefore lead to different knockdown levels. This dependency of gene silencing effectiveness on RNAi expression format was also consistent with a previous report[26]. Therefore, it is critical to confirm that the improved trait is still retained upon the integration of a selected RNAi cassette. It is also possible to avoid the changes in expression contexts by direct integration of the RNAi library into the host genome for library creation, which should further speed up and automate the entire process of RAGE (**Chapter 5**). While the transformation efficiency of an episomal plasmid can easily achieve a complete coverage of the entire genome, the integration efficiency needs to be greatly improved to create a comprehensive library, possibly through the introduction of double-stranded breaks (DSBs) on a genome[52]. Recent development in creating site-specific DSBs by engineered nucleases (zinc-finger nucleases or ZFNs[53], transcription-activator like effector nucleases or TALENs[54, 55], and clustered regularly interspaced short palindromic repeats (CRISPR) or CRISPR-associated nucleases[56, 57]) may facilitate the automation of RAGE by direct integration (**Chapter 5**).

For the engineered strains obtained by RAGE, further investigation of the mechanisms underlying the superior HAc tolerance will provide invaluable knowledge about the target traits. Based on the known functions of the selected knockdown gene targets (**Table 3.5**), it is noteworthy that *ypk3* and *ptc6* are involved in the Target of rapamycin (TOR) pathway, which is responsible for the signalling of acetic acid-induced apoptosis[58]. The gene *rnt1* can mediate selective mRNA degradation and thus regulate the cell wall stress response[59], which is experienced by HAc-challenged yeast cells[39]. The gene *ypk3* might also play a role in the cell wall integrity as a homolog of *ypk1*[60], and the deletion of *ypk1* can improve acid tolerance of the yeast cells[61]. Though the function of *ypr084w* is unknown, *ypr084w* exhibits negative

interactions with the *slm4* and *spf1* genes[62], which are the determinants of HAc resistance[63]. Reprogramming of tRNA modifications was reported to regulate the stress response of yeast cells[64, 65], and it will be interesting to further investigate how our selected RNAi construct for $tRNA^{Val(\underline{A}AC)}$ plays a role in such process.

It is noted that the knockout mutations of the selected genes failed to improve HAc tolerance as the knockdown mutations did (**Fig. 3.8A**). The phenomenon that the performance of a knockdown mutant strain was better than both the wild-type and knockout strains has been observed previously in *S. cerevisiae*[66]. For some essential genes whose knockout mutations are lethal, the reduction-of-function mutants exhibit growth advantages compared to the wild-type strain[66]. We speculate that the underlying reason may be that the knockout mutations of such genes may have dual roles, which result in not only improved tolerance to a certain chemical but also impaired general fitness. Indeed, the knockout strains of the selected gene targets in the first round of RAGE grew at reduced rates compared to the CAD strain without HAc stress (**Fig. 3.7B**), indicating the complete disruption of the normal functions of these genes may affect the general fitness of the yeast cells. Therefore an optimal knockdown level may exist to balance the trade-off between general fitness and inhibitor tolerance. The optimal knockdown level may be determined by fine-tuning the knockdown level of a selected target, which is possible by switching among different RNAi reagent forms (anti-sense RNAs or dsRNAs) and varying the lengths and positions of the RNAi cassettes. These strategies successfully enabled different levels of suppression of both the *gfp* gene (**Fig. 3.3**) and the *ptc6* gene (**Fig. 3.7C** and **Fig. 3.7D**).

Here we show the *de novo* creation of an artificial RNAi regulatory mechanism tailored for the host and target trait(s) in an organism that lacks a native RNAi machinery. RAGE employs a genomic DNA/cDNA-derived library[27] and thus does not require genome sequence information

or a pre-constructed gene-knockout library. Such strategy should be widely applicable in any host of interest with basic genetic tools and a functional RNAi pathway (native or engineered). Although RNAi is only conserved in eukaryote, a recent method based on synthetic small regulatory RNAs (sRNAs) was developed to modulate the expression of up to 122 chromosomal genes in *E. coli*[67]. Thus it is also possible to extend the application of RAGE in bacteria by accumulating beneficial sRNAs in the *E. coli* genome to continuously improve a target trait. Given the versatile tools and broad applications available for both RNA interference and directed evolution, we envision RAGE will become a powerful genome-scale engineering tool for studies in biology, medicine and biotechnology.

3.4. Methods and Materials

3.4.1. Strains, media, and cultivation conditions

S. cerevisiae strain CEN.PK2-1c (*MAT**a** ura3-52 trp1-289 leu2-3,112 his3Δ1 MAL2-8C SUC2*) was purchased from EUROSCARF (Frankfurt, Germany). Zymo 5α Z-competent *E. coli* (Zymo Research, Irvine, CA) and NEB 5α Electrocompetent *E. coli* (New England Biolabs, Ipswich, MA) were used for plasmid amplification and library construction, respectively. *S. castellii* strain was obtained from the ARS culture collection (NRRL number Y-12630) (Peoria, IL). *S. cerevisiae* strains were cultivated in either synthetic dropout medium (0.17% Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate and 0.083% amino acid drop out mix, 0.01% adenine hemisulfate and 2% glucose) or YPAD medium (1% yeast extract, 2% peptone, 0.01% adenine hemisulfate and 2% glucose). 10% (v/v) filtered acetic acid (HAc) solution stock was added into the above medium to make stressed medium.

Before mixing, both the medium and HAc stock were adjusted to pH = 4.5. *S. cerevisiae* strains were cultured at 30 °C and with 250 rpm agitation in baffled shake-flasks for aerobic growth, and at 30 °C and 100 rpm in un-baffled shake-flasks for fermentation. *E. coli* strains were cultured at 37 °C and 250 rpm in Luria broth (LB) medium (Fisher Scientific, Pittsburgh, PA) with the supplement of 100 μg/mL ampicillin. *S. castellii* strain was cultured in YPAD medium at 30 °C and 250 rpm. All chemicals were purchased through Sigma-Aldrich or Fisher Scientific.

3.4.2. DNA manipulation

Plasmid cloning was mostly done by In-fusion HD cloning Kit (Clontech Laboratories, Mountain View, CA) following the manufacturer's instructions or by the DNA assembler method [68]. The complete list of the plasmids, primers and strains in this study is summarized in **Tables 3.6-3.8**. For DNA manipulations, yeast plasmids were isolated using a Zymoprep II yeast plasmid isolation kit (Zymo Research, Irvine, CA) and transferred into *E. coli* for amplification. QIAprep Spin Plasmid Mini-prep Kits (Qiagen, Valencia, CA) were employed to prepare plasmid DNA from *E. coli*. Yeast genomic DNA was isolated by Wizard Genomic DNA Purification Kit (Promega, Madison, WI). All enzymes used for recombinant DNA cloning were from New England Biolabs unless otherwise noted. The products of PCR, digestion and ligation reactions were purified by QIAquick PCR Purification and Gel Extraction Kits (Qiagen, Valencia, CA).

3.4.3. Reconstitution of the RNAi machinery in *S. cerevisiae*

The genomic DNA of *S. castellii* strain was isolated and used as template to clone the *ago1* and *dcr1* genes. An expression cassette in which the *ago1* and *dcr1* genes were driven by constitutive promoters P_{TEF1} and P_{TPI1}, respectively, was assembled into a delta-integration vector by the DNA assembler method. The integration copy number of the RNAi pathway was analyzed by quantitative PCR (qPCR) with LightCycler® 480 SYBR Green I Master (Roche, Indianapolis, IN) following the manufacturer's instructions. To test whether the RNAi machinery is functional, a reporter system based on green fluorescent protein (GFP) was devised. The GFP fluorescence of the engineered strains was analyzed by LSR II Flow Cytometer (BD, Franklin Lakes, NJ).

3.4.4. Construction of a genome-wide RNAi library

Two strong constitutive promoters, P_{TEF1} and P_{TPI1}, were cloned in opposite directions to drive the *in vivo* synthesis of a dsRNA molecule[27]. A *Bam*HI restriction site is engineered between the two promoters to facilitate the insertion of genomic DNA fragments generated by complete *Sau3*AI digestion. The expression cassette was placed on a single-copy plasmid pRS416. The ligation products of the genomic DNA fragments and the linearized vector were transformed into *E. coli* cells by electroporation to create a pooled plasmid library. Specifically, the genomic DNA of yeast strain CEN.PK2-1c was isolated. In a 50 μL reaction, a titration of 0.4-4 U *Sau3*AI enzyme was added to digest 50 μg genomic DNA for 15 min at 37 °C. After heat inactivation at 65 °C for 20 min, the digestion mixtures were column purified, and 0.5 μL of the purified DNA was loaded along with the intact genomic DNA on a 0.7% agarose gel. After

staining with ethidium bromide and visualizing under UV light, the digestion reaction with 4 U *Sau3*AI was considered complete and used for the following experiments.

The ligation was performed by T4 ligase with 300 ng genomic DNA fragments and 100 ng pRS416-TTrc plasmid linearized by *Bam*HI. The single-copy plasmid pRS416 was chosen as the vector to ensure that each yeast cell only carried one RNAi cassette in the library. The ligation product was *n*-butanol precipitated and resuspended in 10 μL ddH_2O, which was transformed with 100 μL NEB 5-alpha Electrocompetent *E. coli* by electroporation. The *E. coli* cells were immediately transferred into 1 mL SOC medium. After recovered for 1 hr, all the cells were used to inoculate 25 mL LB medium plus 100 μg/mL ampicillin (Amp) and allowed to grow until saturation. One hundredth of the cells was plated on an LB+Amp plate to estimate the transformation efficiency. A library size of 4.22×10^5 was obtained, while the control reaction with only the digested plasmid gave only 1.4×10^3 transformants. The plasmid library was isolated from the overnight *E. coli* culture.

3.4.5. Construction of a yeast knockdown library and screening

In the CAD strain, the standard LiAc/ssDNA/PEG protocol[69] was used to transform 20 μg RNAi library plasmids or the control plasmid. Twenty micro-gram of the library DNA was able to achieve a library size more than 5×10^5 to ensure a complete coverage of the yeast genome. Following the transformation, the yeast cells were recovered in 1 mL YPAD medium for 4 hr, and then washed with ddH_2O. For *yku70*Δ suppressor screening, the transformants were spread onto 15 mm diameter petri-dish plates of solid SC-U medium. The amount of cells was adjusted so that each plate would form about 10^4 colonies at a permissive temperature 30 °C and about 10^3 colonies when challenged with 37 °C. The library plates and the control plates were

incubated at 37 °C for 3~4 days. Ninety-three colonies whose sizes were bigger than the largest colonies on the control plates were picked from the library plates. The growth performances of the selected colonies and the control strain were compared in SC-U medium at 37 °C. The initial OD_{600} for all the strains were adjusted to 0.2, and the growth was monitored at 4, 8, 12 and 24 hrs time intervals. The RNAi plasmids from the top strains whose growth behaviors were considerably better than the control strain were isolated and amplified by *E. coli*. The selected plasmids were then individually re-transformed, of which four were able to retain the enhanced HAc tolerance in a fresh background with three biological replicates. The confirmed plasmids were sent for DNA sequencing analysis. The BLAST search was used to identify the sources of the inserts in the selected RNAi plasmids. For each new-identified gene target, a designed RNAi mutant and a knockout mutant were constructed for further analysis. A similar procedure was employed for HAc tolerance screening, except that 0.5% (v/v) HAc was supplemented into the growth medium and the cells was cultured at 30 °C.

3.4.6 Analysis of the newly discovered RNAi targets

To gain a deeper understanding of the selected RNAi targets, both the designed RNAi and the deletion mutants were created and tested for HAc tolerance. The designed RNAi plasmids were constructed by cloning a different region from each target transcript other than the selected plasmids into the pRS416-TTrc plasmid (**Tables 3.2** and **3.4**). The selected plasmids were named as pRS416-TTrc-*gene name*-Si, while the designed plasmids were named as pRS416-TTrc-*gene name*-Di. To perform gene deletion, the open reading frames (ORFs) of target genes were disrupted by inserting the LoxP-*ura3*-LoxP cassette, which was PCR-amplified from plasmid pUG72. The primers were designed in such a way that the *ura3* cassette was flanked by

40 bp sequences that were homologous to the upstream and downstream regions of the target ORFs. The same procedure was used to knockout the *nap1* and *fps1* genes that were known to improve HAc tolerance in BY4741 strain. The strains containing selected RNAi, designed RNAi, deletion and the control plasmid pRS416-TTrc were compared in parallel for the aerobic growth capacity in the presence of HAc stress. Three biological replicates of each strain were inoculated in 3 mL SC-U to grow until saturation. Then 50 μL culture was used to inoculate 3 mL fresh SC-U to synchronize the growth phase. After 20 hrs, the stationary-phase cells were transferred into 4 mL SC-U medium with 0.5% (v/v) HAc. The amount of inoculation was adjusted so that the starting OD_{600} was 0.2. The cell density was monitored at 4, 8, 12, 24 and 36 hr. The normalized OD_{600} values with the control strain as the reference were used to evaluate the HAc tolerance. The same procedure was used to assay the HAc tolerance unless specified otherwise.

3.4.7. Second and third rounds of RAGE

A similar screening procedure was employed in the subsequent rounds as in the first one. For the second round, the four selected RNAi cassettes in the first round were subcloned into the multiple cloning site of pRS403 plasmid and integrated into the *his3* locus of the CAD strain to construct the new parent strains. Then the RNAi plasmid library was transformed into the new parent strains to perform the second round of RAGE. The stress level of 0.6% (v/v) HAc was applied for the screening on the solid medium and the growth quantification in the liquid medium. For the third round of screening, the RNAi cassette for *ypr084w* was subcloned into pRS404 and integrated into the *trp1* locus of CAD_*ptc6i* to create the new parent strain. The stress level of 0.7% was applied. The details about all the selected and designed RNAi cassettes can be found in **Table 3.4**.

3.4.8. Characterization of engineered strains for HAc tolerance

The engineered R1, R2 and R3 strains, together with the control strain, were tested for the biomass accumulation in the synthetic dropout medium SC-U containing 0, 0.5, 0.7, 0.9 and 1.0% (v/v) HAc. Three biological replicates of each strain were inoculated in 3 mL SC-U to grow until saturation. Then 50 μL culture was used to inoculate 3 mL fresh SC-U to synchronize the growth phase. After 20 hrs, the stationary-phase cells were transferred into three culture tubes containing 3 mL SC-U medium and varying concentrations of HAc. The initial OD_{600} was adjusted to 0.01 and 1 mL cell culture was taken periodically at 24, 48 and 72 hrs to measure the cell density. The same procedure was used to compare the R3 strain and the strains containing only one of the three RNAi cassettes from the R3 strain.

The fermentation performance of the R3 strain and the control strain was compared. Three biological replicates of each strain were inoculated in 3 mL SC-U medium in 15 mL round-bottom Falcon tubes to grow until saturation. Then 1 mL culture was transferred into 20 mL SC-U medium in 125 mL baffled shake-flasks. The preparation of the seed culture was performed aerobically (30 °C and 250 rpm). After 20 hrs, the stationary-phase cells were transferred into 50 mL SC-U medium in un-baffled 250 mL shake-flasks. The fermentation was carried out under an oxygen-limited condition (30 °C and 100 rpm). Three levels of HAc stress were applied as 0%, 0.7% and 1.0% (v/v) in the synthetic dropout medium SC-U. 1 mL samples were taken at 0, 4, 8, 12, 24 and 48 hrs for the measurement of cell density and HPLC analysis. An HPX-87H column (BioRad, Hercules, CA) coupled with a refractive index detector (Shimadzu Scientific Instruments, Columbia, MD) was used to separate and analyze the concentrations of glucose, ethanol, and acetate in the broth following the manufacturer's instructions.

3.4.9. Estimation of the gene knockdown efficiency

By knocking-in the GFP gene as a carboxy-terminal fusion reporter, it is possible to quantify the expression levels of the target proteins. All the primers and the process to create a perfect in-frame GFP fusion protein were reported elsewhere[33]. We measured the knockdown efficiency for the four target genes found in the first round of RAGE. The cells were cultured aerobically in SC-His/Ura and subjected to the flow cytometry analysis in their log-phase.

To determine the knockdown efficiency of the other two RNAi cassettes (*ypr084w* and $tRNA^{Val(\underline{A}AC)}$), a semi-quantification assay was employed to measure the level of target RNA molecules[24]. The reason why we did not use the GFP-fusion construct is that the *his3* selection marker used to insert GFP cassette was already used to integrate the RNAi cassettes. The total RNAs from the R3 strain and the control strain were isolated using the RNeasy mini kit (QIAGEN, Valencia, CA). The cDNA synthesis was performed by the Transcriptor First Strand cDNA Synthesis Kit (Roche, Indianapolis, IN). In each reverse-transcription reaction for a certain gene, 5 μg of the RNA was used as template, and the gene-specific primers for the target gene and the internal control gene *act1* were added in the same reaction. For the control samples, only the reverse-transcriptase was omitted. The primers for the cDNA synthesis were named as *gene-name*-cDNA Rev (**Table 3.7**). For the PCR reactions, 1 μL of the RT-reaction was added to a 100 μL reaction. The primers were named as *gene-name*-RT-PCR For/Rev. The PCR reaction was dispensed as 20 μL aliquots and run for the indicated numbers of cycles (16, 20 and 24 for act1, 20, 24 and 28 for *ptc6* and *ypr084w*). Then 4 μL 6X loading dye was mixed with the PCR reaction, and 6 μL of the mixture was loaded onto 1.5% agarose gel for ethidium bromide staining and UV visualization (**Fig. 3.11**). No RT-PCR products were obtained after 32 cycles for the control samples or for the $tRNA^{Val(\underline{A}AC)}$ samples.

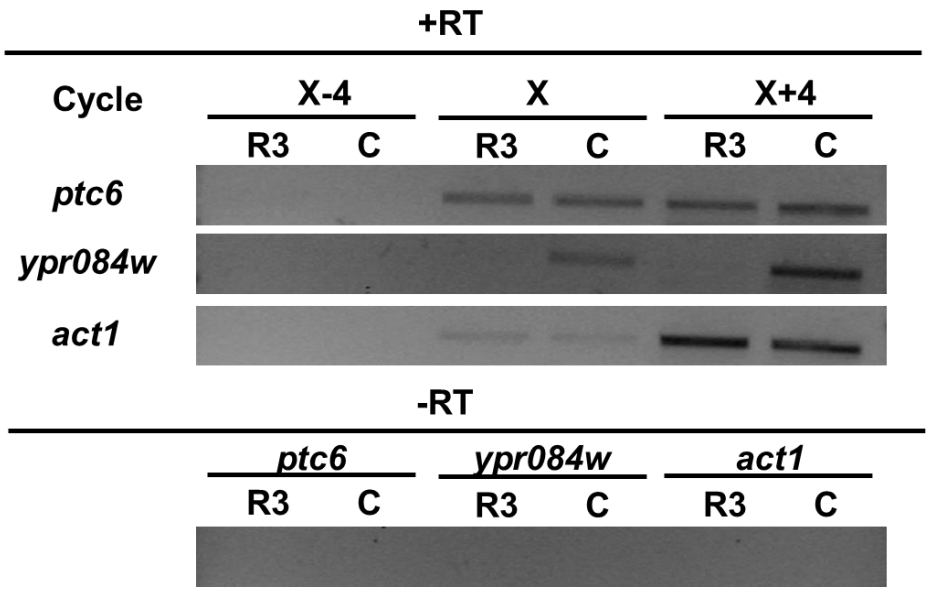

Figure 3.11. Semi-quantification of mRNA levels. RT-PCR products with increasing cycles of amplification were loaded onto 1.5% agarose gel and stained by ethidium bromide. Cycle number: X=20 for the *act1* gene, and X=24 for the *ptc6* and *ypr084w* genes. No RT-PCR product was detected in control samples or for $tRNA^{Val(\underline{A}AC)}$ after 32 cycles.

Table 3.6 List of plasmids used in this chapter

Plasmid	Primers for PCR	Notes
pRS425-TEF1p-*Pme*I-PGK1t	pRS-TEF1p-40bp For/TEF1p-*Pme*I-PGK1t	Helper plasmid to clone genes into the TEF1p-PGK1t cassette; constructed by DNA assembler
	TEF1p-*Pme*I-PGK1t For/PGK1t-pRS-40bp	
pRS425-PGK1t-TPI1p-*Pme*I-GPDt	pRS-PGK1t For/PGK1t-TPI1p Rev	Helper plasmid to clone genes into the TPI1p-GPDt cassette; constructed by DNA assembler
	PGK1t-TPI1p For/TPI1p-*Pme*I-GPDt Rev	
	TPI1p-*Pme*I-GPDt For/GPDt-pRS Rev	
pRS416-GPDtrc-TEF1p-*Bam*HI-TPI1prc-PGK1t (aka pRS416-TTrc)	pRS-GPDtrc For/GPDtrc Rev	Helper plasmid with convergent promoters to produce dsRNA
	GPDtrc-TEF1p For/TEF1p-*Bam*HI-TPI1prc	
	*Bam*HI-TPI1prc For/TPI1prc-PGK1t Rev	
	TPI1prc-PGK1t For/PGK1t-pRS-40bp Rev	
pRS425-TEF1p-GFP-PGK1t	TEF1p-GFP-1 For/GFP-PGK1t Rev	Construction of expression cassette for GFP
pRS425-TEF1p-GFPrc-PGK1t	TEF1p-GFPrc For/GFPrc-PGK1t Rev	Construction of anti-sense expression cassette for GFP
pRS405-TEF1p-GFP-PGK1t	pRS-TEF1p For/PGK1t-pRS Rev	Integration of GFP expression cassette into *leu2* site
pRS416-TEF1p-GFPrc-PGK1t	pRS-TEF1p For/PGK1t-pRS Rev	Anti-sense expression cassette for GFP on a single
pRS416-TTrc-GFP-223-631	TEF1p-GFP-223 For/GFP-631-TPI1prc Rev	To express dsRNA derived from 223-631bp, 1-180bp, 1-360bp, 1-540bp and 1-717bp of GFP by convergent promoters
pRS416-TTrc-GFP-1-180	TEF1p-GFP-001 For/GFP-180-TPI1prc Rev	
pRS416-TTrc-GFP-1-360	TEF1p-GFP-001 For/GFP-360-TPI1prc Rev	
pRS416-TTrc-GFP-1-540	TEF1p-GFP-001 For/GFP-540-TPI1prc Rev	
pRS416-TTrc-GFP-1-717	TEF1p-GFP-001 For/GFP-717-TPI1prc Rev	
pRS425-TEF1p-AGO1-PGK1t	TEF1p-AGO-1 For/AGO-1 Rev	Helper plasmid to clone *ago1* into TEF1p-PGK1t cassette
	AGO-2 For/AGO-2-PGK1t Rev	
pRS425-PGK1t-TPI1p-DCR1-	TPI1p-DCR1 For/DCR1-GPD1t Rev	Helper plasmid to clone *dcr1*into TPI1p-GPDt cassette
pRS-delta-KanMX-LoxP-TEF1p-AGO1-PGK1t-TPI1p-DCR1-	pRS For/PGK1t Rev	Integration of *S. castellii* RNAi pathway into delta-site, constructed by DNA assembler
	PGK1t For/pRS Rev	
pRS403-RNAi-cassette	pRS-TEF1p F/PGK1t-pRS Rev	Integration of selected RNAi cassettes into the *his3*
pRS404-RNAi-cassette		Integration of selected RNAi cassettes into the *trp1*
pRS405-RNAi-cassette		Integration of selected RNAi cassettes into the *leu2*
pRS406-RNAi-cassette		Integration of GFP RNAi cassettes into the *ura3* locus

Table 3.7. List of primers used in this chapter

Primer name	Primer sequence (5'→3')
pRS-TEF1p-40bp For	TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACAGCAACAGGCGCGTTGGAC
TEF1p-*Pme*I-PGK1t Rev	AATTGATCTATCGATTTCAATTCAATTCAATGTTTAAACTTTGTAATTAAAACTTAG
TEF1p-*Pme*I-PGK1t For	CTAAGTTTTAATTACAAAGTTTAAACATTGAATTGAATTGAAATCGATAGATCAATT
PGK1t-pRS Rev	GATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACCAGGAAGAATACACTATAC
pRS-PGK1t-40bp For	GTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACATTGAATTGAATTGAAATCG
PGK1t-TPI1p Rev	TAATCTTCCACCAACCTGATGGGTTCCTAGATATACAGGAAGAATACACTATAC
PGK1t-TPI1p For	GTATAGTGTATTCTTCCTGTATATCTAGGAACCCATCAGGTTGGTGGAAGATTA
TPI1p-*Pme*I-GPDt Rev	ATTTAAATGCAAGATTTAAAGTAAATTCACGTTTAAACTTTTAGTTTATGTATGTG
TPI1p-*Pme*I-GPDt For	CACATACATAAACTAAAAGTTTAAACGTGAATTTACTTTAAATCTTGCATTTAAAT
GPDt-pRS Rev	GATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACGGAATCTGTGTATATTACTG
pRS-GPDtrc For	CGTAATACGACTCACGGAATCTGTGTATATTAC
GPDtrc Rev	GTGAATTTACTTTAAATC
GPDtrc-TEF1p For	TTAAAGTAAATTCACAGCAACAGGCGCGTTGGAC
TEF1p-*Bam*HI-TPI1prc Rev	ATACATAAACTAAAAGGATCCTTTGTAATTAAAACTTAGATTAG
*Bam*HI-TPI1prc For	TTTTAGTTTATGTATGTG
TPI1prc-PGK1t Rev	TCAATTCAATTCAATTATATCTAGGAACCCATC
TPI1prc-PGK1t For	GATGGGTTCCTAGATATAATTGAATTGAATTGA
pRS-TEF1p For	GGCGAATTGGGTACCAGCAACAGGCGCGTTGGAC
PGK1t-pRS Rev	AAAGCTGGAGCTCCACAGGAAGAATACACTATAC
TEF1p-GFP-1 For	ATAGCAATCTAATCTAAGTTTTAATTACAAAATGAGTAAAGGAGAAGAAC
GFP-PGK1t Rev	AATTGATCTATCGATTTCAATTCAATTCAATCTATTTGTATAGTTCATCC
TEF1p-rcGFP For	ATAGCAATCTAATCTAAGTTTTAATTACAAACTATTTGTATAGTTCATCC
rcGFP-PGK1t Rev	AATTGATCTATCGATTTCAATTCAATTCAATATGAGTAAAGGAGAAGAAC
TEF1p-GFP-226 For	AGTTTTAATTACAAAGATCATATGAAACGGCATGAC
GFP-631-TPIprc Rev	ATACATAAACTAAAAGATCTTTCGAAAGGGCAGATTG
TEF1p-GFP-001 For	AGTTTTAATTACAAAATGAGTAAAGGAGAAGAAC
GFP-180-TPIrc Rev	ATACATAAACTAAAAAAGTGTTGGCCATGGAAC
GFP-360-TPIrc Rev	ATACATAAACTAAAAAACAAGGGTATCACCTTC
GFP-540-TPIrc Rev	ATACATAAACTAAAAGTCTGCTAGTTGAACGCTTCC
GFP-717-TPIrc Rev	ATACATAAACTAAAACTATTTGTATAGTTCATCC
TEF1p-AGO-1 For	ATAGCAATCTAATCTAAGTTTTAATTACAAAATGTCATCCAATTCGGAGG
AGO-1 Rev	CCTTTTCACCTGGTTTTGG
AGO-2 For	CCAAAACCAGGTGAAAAGG
AGO-2-PGK1t Rev	AATTGATCTATCGATTTCAATTCAATTCAATTCATATGTAGTACATGATG
TPI1p-DCR1 For	AACTACAAAAAACACATACATAAACTAAAAATGAATAGAGAAAAAAGCGCCGATC
DCR1-GPD1t Rev	ATTTAAATGCAAGATTTAAAGTAAATTCACTCACAGATTGTTGCAATGCCTC
pRS For	CGAGGTGCCGTAAAGCAC
PGK1t Rev	CAGGAAGAATACACTATAC
PGK1t For	ATTGAATTGAATTGAAATCG
pRS Rev	GCTCACATGTTCTTTCC
qAgo For	TGCCGTTCCACAGAGTAATTC
qAgo Rev	TCACCATTGAAGGAGATACGATC
qDrc For	AAAATGCACCTATCCCTATCCC
qDcr Rev	TGGGACTGAGACTGAGACTG
S TEF1p For	TTTTACTTCTTGCTCATTAG
S TPI1p For	TTTTTGTTTGTATTCTTTTC

Table 3.7. List of primers used in this chapter (continued)

Primer name	Primer sequence (5'→3')
ypr084w-cDNA Rev	GAGAGTTGTAAATTGAACTTG
ypr084w-RT-PCR For	CGAGGTTCATTATCCAAC
ypr084w-RT-PCR Rev	CTGTGTTCATCACATCATC
ptc6-cDNA Rev	ATTCATTAGTTTGGTTTCTC
ptc6-RT-PCR For	CTGCAACTCGAAACAAGAAC
ptc6-RT-PCR Rev	TAGAGGTGATGAGGTCAAC
val-tRNA-cDNA Rev	TTTCGCCCAGGATCGAAC
val-tRNA-RT-PCR For	GGTTTCGTGGTCTAG
val-tRNA-RT-PCR Rev	GAACTGGGGACGTTC
GFP-cDNA Rev	TGTGGTCTCTCTTTTCGTTGG
GFP-RT-PCR For	TTTCACTGGAGTTGTCCCAAT
GFP-RT-PCR Rev	GAAAGGGCAGATTGTGTGG
act1-cDNA Rev	TCAAAGAAGCCAAGATAGAACCA
act1-RT-PCR For	ACGTTGGTGATGAAGCTCAA
act1-RT-PCR Rev	ATACCTGGGAACATGGTGGT
TEF1p-*rnt1*-Di For	AGTTTTAATTACAAAGTTAAAAGGGAAAAGCGAGA
rnt1-Di-TPIprc Rev	AAGATTTAGCGATTAGGGCCTTTTAGTTTATGTAT
TEF1p-*ymr1*-Di For	AGTTTTAATTACAAAACGAAAGATTTTTAAGAAGA
ymr1-Di-TPIprc Rev	AAATTTCATACAGAGAAAAATTTTAGTTTATGTAT
TEF1p-*ptc6*-Di For	AGTTTTAATTACAAAATTGTCACGCAGGTGGGCGA
ptc6-Di-TPIprc Rev	TTTTGAATAATTTCGCCAACTTTTAGTTTATGTAT
TEF1p-*ypk3*-Di For	AGTTTTAATTACAAAGCAATTACGGAAAGCAGAAA
ypk3-Di-TPIprc Rev	CATCCAAACATTGTCAAGTTTTTTAGTTTATGTAT

Table 3.8. List of strains used in this chapter

Strain	Genotype	Reference
CEN.PK2-1c	*MAT**a** ura3-52 trp1-289 leu2-3,112 his3Δ1 MAL2-8C SUC2*	EUROSCARF
CAD	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1*	This study
CAD_GFP	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 leu2::* P_{TEF1}*-GFP*	This study
CEN_GFP	*MAT**a** leu2::* P_{TEF1}*-GFP*	This study
CAD_*yku70*Δ	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 yku70*Δ*::ura3*	This study
CAD_*rnt1i*	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 his3::TTrc-rnt1-Si*	This study
CAD_*ymr1i*	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 his3::TTrc-ymr1-Si*	This study
CAD_*ptc6i*	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 his3::TTrc-ptc6-Si*	This study
CAD_*ypk3i*	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 his3::TTrc-ypk3-Si*	This study
CAD_*ymr1*Δ	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 ymr1*Δ*::ura3*	This study
CAD_*ptc6*Δ	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 1ptc6*Δ*::ura3*	This study
CAD_*ypk3*Δ	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 ypk3*Δ*::ura3*	This study
CAD_*ptc6i_ypr084wi*	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 his3::TTrc-ptc6-Si trp1::TTrc-ypr084w*	This study
CAD_*ptc6i_ypr084wi_tRNAi*	*MAT**a** δ::*P_{TEF1}*-ago1*-P_{TPI1}*-dcr1 his3::TTrc-ptc6-Si trp1::TTrc-ypr084w leu2:: TTrc-* $tRNA^{Val(\underline{A}AC)}$	This study
R1 (in Fig. 3.9)	CAD_*ptc6i* strain with pRS416-TTrc	This study
R2 (in Fig. 3.9)	CAD_*ptc6i_ypr084wi* strain with pRS416-TTrc	This study
R3 (in Fig. 3.9)	CAD_*ptc6i_ypr084wi_tRNAi* strain with pRS416-TTrc	This study

3.5. References

1. Almeida, J.R.M. et al. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *Curr Opin Biotechnol* **82**, 340-349 (2007).
2. Santos, C.N.S. & Stephanopoulos, G. Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr Opin Chem Bio* **12**, 168-176 (2008).
3. Woodruff, L.B.A. & Gill, R.T. Engineering genomes in multiplex. *Curr Opin Biotechnol* **22**, 576-583 (2011).
4. Cakar, Z.P., Turanli-Yildiz, B., Alkim, C. & Yilmaz, U. Evolutionary engineering of *Saccharomyces cerevisiae* for improved industrially important properties. *FEMS Yeast Res* **12**, 171-182 (2012).
5. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).
6. Alper, H., Moxley, J., Nevoigt, E., Fink, G.R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565-1568 (2006).
7. Gibson, D.G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215-1220 (2008).
8. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B.A. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**, 856-U138 (2010).
9. Nevoigt, E. Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **72**, 379-412 (2008).

10. Hong, K.-K. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol Life Sci* **69**, 2671-2690 (2012).

11. Winzeler, E.A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).

12. Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391 (2002).

13. Tong, A.H.Y. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364-2368 (2001).

14. Forsburg, S.L. The art and design of genetic screens: yeast. *Nat Rev Genet* **2**, 659-668 (2001).

15. Scherens, B. & Goffeau, A. The uses of genome-wide yeast mutant collections. *Genome Biol* **5**, 229 (2004).

16. van Dijken, J.P. et al. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* **26**, 706-714 (2000).

17. Arnold, F.H. Design by directed evolution. *Acct Chem Res* **31**, 125-131 (1998).

18. Cobb, R.E., Si, T. & Zhao, H. Directed evolution: an evolving and enabling synthetic biology tool. *Curr Opin Chem Bio* **16**, 285-291 (2012).

19. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-437 (2004).

20. Echeverri, C.J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7**, 373-384 (2006).

21. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811 (1998).

22. Hannon, G.J. RNA interference. *Nature* **418**, 244-251 (2002).

23. Boutros, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nat Rev Genet* **9**, 554-566 (2008).

24. Drinnenberg, I.A. et al. RNAi in budding yeast. *Science* **326**, 544-550 (2009).

25. Suk, K. et al. Reconstitution of human RNA interference in budding yeast. *Nucleic Acids Res* **39**, e43 (2011).

26. Crook, N.C., Schmitz, A.C. & Alper, H.S. Optimization of a Yeast RNA Interference System for Controlling Gene Expression and Enabling Rapid Metabolic Engineering. *ACS Synth Biol* (2013).

27. Cheng, X. & Jian, R. Construction and application of random dsRNA interference library for functional genetic screens in embryonic stem cells. *Methods Mol Biol* **650**, 65-74 (2010).

28. Lander, E.S. & Waterman, M.S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231-239 (1988).

29. Wellinger, R.J. & Zakian, V.A. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics* **191**, 1073-1105 (2012).

30. Kanoh, J. & Ishikawa, F. Composition and conservation of the telomeric complex. *Cell Mol Life Sci* **60**, 2295-2302 (2003).

31. Addinall, S.G. et al. Quantitative fitness analysis shows that NMD proteins and many other protein complexes suppress or enhance distinct telomere cap defects. *PLoS Genet* **7**, e1001362 (2011).

32. Vandre, C. in Cell and Developmental Biology, Vol. PhD (University of Illinois at Urbana-Champaign, Urbana; 2008).

33. Huh, W.K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691 (2003).

34. Jackson, A.L. et al. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**, 635-637 (2003).

35. Addinall, S.G. et al. A Genomewide suppressor and enhancer analysis of cdc13-1 reveals varied cellular processes influencing telomere capping in *Saccharomyces cerevisiae*. *Genetics* **180**, 2251-2266 (2008).

36. Krogan, N.J. et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637-643 (2006).

37. Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R. & Monot, F. New improvements for lignocellulosic ethanol. *Curr Opin Biotechnol* **20**, 372-380 (2009).

38. Olsson, L. & HahnHagerdal, B. Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme Microb Technol* **18**, 312-331 (1996).

39. Mira, N.P., Teixeira, M.C. & Sa-Correia, I. Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: a genome-wide view. *OMICS* **14**, 525-540 (2010).

40. Keck, K.M. & Pemberton, L.F. Interaction with the histone chaperone Vps75 promotes nuclear localization and HAT activity of Rtt109 in vivo. *Traffic* **12**, 826-839 (2011).

41. Mollapour, M. & Piper, P.W. Hog1 mitogen-activated protein kinase phosphorylation targets the yeast Fps1 aquaglyceroporin for endocytosis, thereby rendering cells resistant to acetic acid. *Mol Cell Bio* **27**, 6446-6456 (2007).

42. Mira, N., Palma, M., Guerreiro, J. & Sa-Correia, I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb Cell Fact* **9**, 79 (2010).

43. Drinnenberg, I.A., Fink, G.R. & Bartel, D.P. Compatibility with Killer explains the rise of RNAi-deficient fungi. *Science* **333**, 1592-1592 (2011).

44. Hebert, C.G., Valdes, J.J. & Bentley, W.E. Beyond silencing - engineering applications of RNA interference and antisense technology for altering cellular phenotype. *Curr Opin Biotechnol* **19**, 500-505 (2008).

45. Dorsett, Y. & Tuschl, T. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* **3**, 318-329 (2004).

46. Le Borgne, S. in Recombinant Gene Expression: Reviews and Protocols, Third Edition, Vol. 824. (ed. A. Lorence) 451-465 (Humana Press Inc, Totowa; 2012).

47. Birmingham, A. et al. 3 ' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods* **3**, 199-204 (2006).

48. Ma, Y., Creanga, A., Lum, L. & Beachy, P.A. Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature* **443**, 359-363 (2006).

49. Bumcrot, D., Manoharan, M., Koteliansky, V. & Sah, D.W.Y. RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat Chem Biol* **2**, 711-719 (2006).

50. Shendure, J. & Aiden, E.L. The expanding scope of DNA sequencing. *Nat Biotechnol* **30**, 1084-1094 (2012).

51. Horn, T., Sandmann, T. & Boutros, M. Design and evaluation of genome-wide libraries for RNA interference screens. *Genome Biol* **11** (2010).

52. Storici, F., Durham, C.L., Gordenin, D.A. & Resnick, M.A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci U S A* **100**, 14994-14999 (2003).

53. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**, 636-646 (2010).

54. Sun, N., Liang, J., Abil, Z. & Zhao, H.M. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst* **8**, 1255-1263 (2012).

55. Joung, J.K. & Sander, J.D. INNOVATION TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**, 49-55 (2013).

56. Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819-823 (2013).

57. DiCarlo, J.E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* **41**, 4336-4343 (2013).

58. Almeida, B. et al. Yeast protein expression profile during acetic acid-induced apoptosis indicates causal involvement of the TOR pathway. *Proteomics* **9**, 720-732 (2009).

59. Catala, M., Aksouh, L. & Abou Elela, S. RNA-dependent regulation of the cell wall stress response. *Nucleic Acids Res* **40**, 7507-7517 (2012).

60. Schmelzle, T., Helliwell, S.B. & Hall, M.N. Yeast protein kinases and the RHO1 exchange factor TUS1 are novel components of the cell integrity pathway in yeast. *Mol Cell Bio* **22**, 1329-1339 (2002).

61. Kawahata, M., Masaki, K., Fujii, T. & Iefuji, H. Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in

Saccharomyces cerevisiae cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Res* **6**, 924-936 (2006).

62. Costanzo, M. et al. The Genetic Landscape of a Cell. *Science* **327**, 425-431 (2010).

63. Mira, N.P., Palma, M., Guerreiro, J.F. & Sa-Correia, I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb Cell Fact* **9**, 79 (2010).

64. Chan, C.T.Y. et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat Commun* **3**, 937 (2012).

65. Preston, M.A., D'Silva, S., Kon, Y. & Phizicky, E.M. tRNAHis 5-methylcytidine levels increase in response to several growth arrest conditions in *Saccharomyces cerevisiae*. *RNA* **19**, 1-14 (2012).

66. Breslow, D.K. et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* **5**, 711-718 (2008).

67. Na, D. et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat Biotechnol* **31**, 170-174 (2013).

68. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**, e16 (2009).

69. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protocols* **2**, 31-34 (2007).

Chapter 4 RAGE 2.0: Constructing a Comprehensive Yeast Library

4.1. Introduction

4.1.1. Comprehensive genome-wide library

Genome-wide strain libraries are powerful tools to comprehensively investigate the impact of individual genetic modification on a given phenotype. Traditionally, it needs to create overexpression, knockout, and knockdown libraries separately, as the techniques for constructing these three types of libraries are fundamentally different. Overexpression libraries are often created on an episomal vector, on which either genomic fragments[1] or all the open reading frames (ORFs) under the control of a promoter[2] can be cloned. Genome-wide knockout libraries requires genome editing to disrupt the ORFs, through either transposon mutagenesis optimized for unbiased integration of an antibiotic marker cassette into the entire genome[3, 4], or homologous recombination to replace the ORFs of non-essential genes with selection markers[5-7]. Knockdown screen can be achieved by both plasmid libraries and genome editing. Plasmid-born regulatory non-coding RNAs (ncRNA) libraries can be introduced for genome-wide gene silencing, as exemplified by anti-sense RNA screen in bacteria[8-10] and RNAi screen in eukaryotes[11-13]. For genome editing, insertion of an antibiotic marker into the terminator region destabilizes the target mRNAs, and this method was used to make knockdown modifications of the essential genes in *S. cerevisiae*[14]. Lacking of methods to create comprehensive genetic library complicates the genotyping effort, as the screen process, which is often resource-intensive and time-consuming, needs to be repeated for different libraries.

Recently, the trackable multiplex recombineering (TRMR) method has been developed to create comprehensive *E. coli* genetic libraries including both overexpression and knockdown modifications[15]. Two kinds of synthetic cassettes were designed for promoter replacement: the 'up' cassette containing a strong promoter, and the 'down' cassette containing an inert sequence to replace the native RBS. Through recombineering, these synthetic cassettes were incorporated in front of every gene in *E. coli*, which led to either increased or decreased expression of a target gene[15]. Nevertheless, it is very difficult to apply TRMR in *S. cerevisiae*, which lacks efficient mechanism for oligonucleotide-mediated allele replacement. On the other hand, as plasmid transformation in *S. cerevisiae* is simple and efficient, we speculate that combination of plasmid-borne overexpression and knockdown libraries should be a more feasible route to create a comprehensive genetic library.

4.1.2. Construction of a normalized, full-length cDNA library

To construct an overexpression library, full-length cDNA clones containing complete ORFs are required. However, the majority of cDNA clones obtained by traditional methods are not full-length, mainly due to premature termination of reverse transcription during the first strand cDNA synthesis[16]. This problem of 5' truncated cDNA synthesis is especially severe for long mRNAs that contain abundant secondary structures[17]. The Switching Mechanism At 5' end of RNA Template (SMART) method was recently developed to enrich for full-length cDNAs, taking advantage of some unique features of the Moloney murine leukemia virus reverse transcriptase (MMLV RTase)[18]. During the first-strand cDNA synthesis, the MMLV RTase acts as a terminal transferase to add a few additional C residues at the 3' end of the cDNA only when it reaches the end of the mRNA template. Such reaction will not happen with incomplete reverse

transcripts. Then an RNA primer ending in three Gs base pairs with the added Cs and serves as an extended template. The MMLV RTase switches templates and continues the cDNA synthesis to incorporate the sequences in the RNA primer with a universal priming site (the SMART anchor). Therefore, only full-length cDNAs containing the SMART anchor at their 5' ends are amplified during subsequent PCR amplification[18]. The PCR amplification step also incorporate unique adaptor sequences at the end of the cDNA molecules to facilitate directional cloning.

Some improvements have been introduced to the SMART technology since its inception. First, size fractionation was performed after the PCR amplification step to remove smaller cDNA fragments that are preferentially amplified by PCR reaction[17]. Second, a normalization step has been incorporated[19], which is necessary due to the substantial variation in the transcript abundance[20] (For the total mRNA mass, 10–20 abundant transcripts comprise about 20%, several hundred medium transcripts comprise 40–60%, and several thousand rare transcripts account for 20–40%). Among many cDNA normalization protocol[19], selective degradation of abundant cDNAs during re-naturation process is considered the most straightforward and effective. As second-order reaction kinetics governs the re-association of denatured DNAs, the double-stranded DNAs (dsDNAs) portion mainly consists of the abundant cDNA individuals, and the remaining single-stranded DNAs (ssDNAs) is therefore relatively equalized[21]. To this end, a duplex-specific nuclease (DSN) from the Kamchatka crab is very useful, because of its strong preference for cleaving dsDNA over ssDNA, irrespective of sequence lengths[22]. This DSN is stable at high temperature (maximal activity observed at 60-65 °C), thus avoiding non-specific degradation of ss cDNAs due to secondary structure formation[21]. Together, the size-fractionation and normalization steps remarkable improve the quality and representativeness of the full-length cDNA libraries[19].

In this chapter, we developed the RAGE2.0 method, where directional cloning of a full-length, normalized cDNA library enabled simultaneous construction of the ORF-overexpression and anti-sense RNA libraries. In the presence of an RNAi pathway, we achieved both genome-wide overexpression and knockdown modifications in *S. cerevisiae* in one step. A series of phenotypes, including protein surface display, substrate utilization, and fuel molecule production, were screened with the RAGE2.0 library in a high-throughput manner. Both gene overexpression and knockdown targets were successfully identified for these phenotypes.

4.2. Results

4.2.1. Directional cloning of a full-length, normalized cDNA library of *S. cerevisiae*

As both overexpression and reduction-of-function screens are important for genotype-phenotype mapping, it is desirable to develop a method to incorporate both types of genetic modifications simultaneously to accelerate the screen process. Inspired by the observation that full-length anti-sense mRNAs can efficiently knockdown a target gene through RNAi in **Chapter 3**, we devised a method for one-step construction of a comprehensive genome-wide library through directional cloning of full-length cDNAs. When a full-length cDNA molecule is inserted in the sense direction under the control of a constitutive promoter, the resultant construct will lead to gene overexpression of the contained ORF. When inserted in the reversed direction, on the other hand, the resultant construct will transcribe the anti-sense mRNA of the target gene, and elicit gene silencing in the presence of RNAi machinery. The directional cloning can be achieved by adding two different adaptor sequences to the ends of a cDNA molecule, and thus

the insertion direction can be controlled by arranging the homologous adaptor sequences in a desired order on destination plasmids.

We first synthesized a full-length cDNA library using a commercial kit based on the SMART technology. Two 15 bp adaptor sequences, 5'-CGGGGTACGATGAGA-3' and 5'-TTGATACCACTGCTT-3', were incorporated at the 5' and 3' ends of the cDNA molecules, respectively (**Fig. 4.1** and **Table 4.6**). The size-fractionation step is included in this kit. After treatment with another commercial kit for normalization, the cDNA library is ready to be cloned into the destination plasmid. For directional cloning, we developed a constitutive expression cassette consisting of a P_{TEF1} promoter and a T_{PGK1} terminator. Two versions of this cassette was constructed (**Fig. 4.1**). While both contain the *ccdB* suicide gene to minimize empty vectors in the library[23], the adaptor sequences flanking the *ccdB* gene in these two versions are arranged in opposite directions. Two *Pme*I sites are incorporated to release the *ccdB* gene and to expose the adaptor sequences upon restriction digestion. The exposed adaptors are designed for homologous recombination with their counterparts on the cDNA molecules through Gibson assembly reaction, resulting in cDNA insertion in either sense or anti-sense direction into the expression cassette. Either the pRS416 or pRS426 plasmid was used as vectors for the expression cassette. To simplify subsequent analysis, the sense and anti-sense plasmid libraries were constructed separately with the Gibson assembly reaction. The reaction mix was then transformed into a *ccdB* sensitive *E. coli* strain by electroporation, and the transformants were recovered on solid medium to minimize uneven amplification of smaller plasmids. The experiments were carefully executed to ensure the number of independent clones in the library to exceed 10^6, representing more than 100-fold redundancy of the 6000 yeast genes.

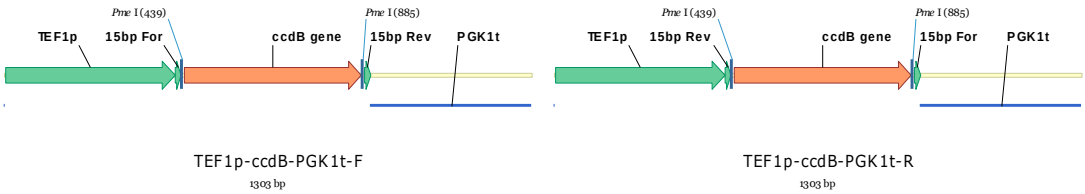

Figure 4.1 Expression cassettes for directional cloning in the RAGE2.0 library. The P_{TEF1} promoter and T_{PGK1} terminator are used to signal the start and end of transcription. Two 15 bp adaptor sequences are incorporated to facilitate directional insertion of cDNAs flanked with the same pair of adaptors. The toxic *ccdB* gene is replaced with the cDNA molecules during the RAG2.0 library construction, and the *E. coli* cells harboring the uncut plasmids with the intact *ccdB* gene cannot survive, hence reducing the presence of empty plasmids in the library.

Ten colonies were randomly picked from each *E. coli* library (overexpression and knockdown), and the isolated plasmids were analyzed by DNA sequencing. The results indicated that all the cDNAs were inserted in the expected directions, and most insets (19/20) were full-length cDNAs. The only partial cDNA had the characteristic poly-A tail in an mRNA molecule, indicating a premature transcriptional intermediate.

Table 4.1 Inset information from twenty random plasmids in the RAGE2.0 library

Overexpression library		Anti-sense library	
Gene name	ORF length (bp)	Gene name	ORF length (bp)
CBK1	2271	*RPS4A*	786
OLE1	1533*	*TOS6*	309
RPS8B	604	*INH1*	258
HSC82	2118	*KAR2*	2049
SCD6	1050	*GPM1*	744
ATP17	306	*SPI1*	447
UFD1	1086	*SSA4*	1929
HSP26	645	*RIB4*	510
LSP1	1026	*GIM3*	390
HTB1	396	*PIL1*	1020

* A fragment (1-264bp) of the OLE1 ORF was cloned

4.2.2. High-throughput screening with the RAGE2.0 library

4.2.2.1 Surface display of a cellulase

Consolidated bioprocessing (CBP) is proposed to substantially lower the cost of cellulosic ethanol production, by combining enzyme production, cellulose hydrolysis and sugar fermentation in a single step[24]. To realize CBP, constructing recombinant *S. cerevisiae* strains with cellulolytic capacity is one promising strategy[24, 25]. Specifically, cellulase-displaying yeast cells has drawn intensive attention[26-28]. Compared with secreting free cellulases, anchoring the cellulases on yeast cell surface is considered more efficient by improving local enzyme concentrations and enabling enzyme-enzyme synergy[26-28]. However, the performance of current cellulolytic *S. cerevisiae* strains is far from optimal, and the low cellulase display level is considered the main obstacle[29].

There are essentially two ways to increase protein display level, including protein engineering and host engineering. Although successful in many surface display applications, such as the insulin precursor[30] and single-chain T-cell receptor (scTCR)[31], protein engineering can be problematic to improve cellulase secretion, as deleterious mutations affecting catalytic activity may occur. Instead, engineering the host cells by rational[32, 33] or random[34] approaches are presumably preferred. As an example of the random approach, combining cDNA overexpression libraries with yeast surface display successfully identified several display-enhancing genes conferring up to 8-fold improvement in scTCR secretion[34]. Notably, whereas rational methods have demonstrated the importance of gene deletion in promoting protein secretion[35, 36], reduction/loss-of-function screens have not been reported yet for this purpose. This is probably due to the difficulty to create genome-wide knockout/knockdown libraries *de novo* in a recombinant protein-displaying strain. Therefore, a comprehensive screen is needed to

include knockdown modifications for the identification of yeast mutants with enhanced surface display capacity.

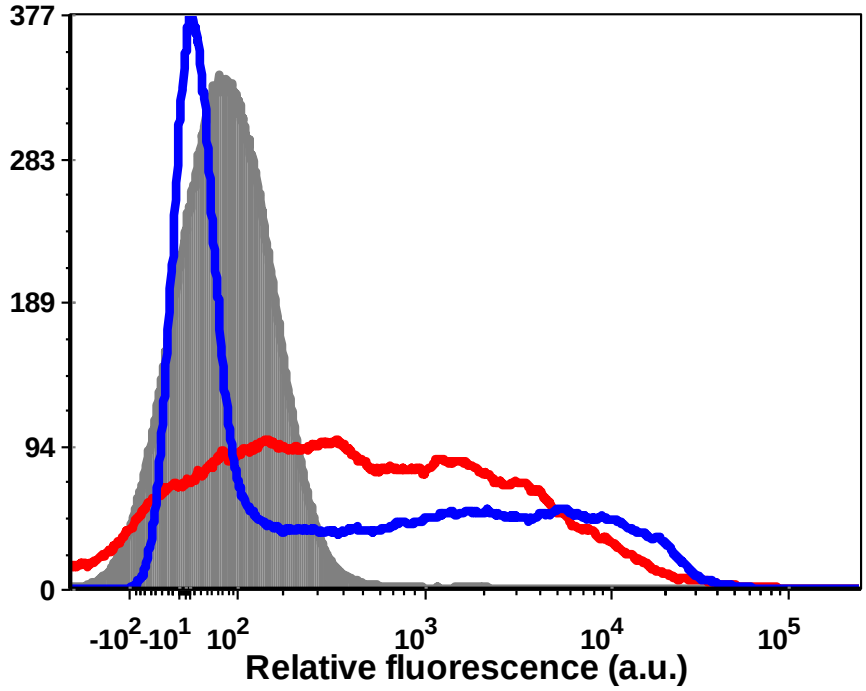

Figure 4.2 Flow cytometry analysis of EGII-displaying yeast cells. Both the plasmid-borne (blue line) and integrated (red line) EGII expression cassettes led to display of EGII on the yeast cell surface indicated by His3 epitope detection. The CAD strain transformed with an empty pRS425 plasmid was used as a negative control (gray shade).

As proof of concept, we performed a high-throughput genetic screen in a cellulase-displaying yeast strain with the RAGE2.0 library. The expression cassette of *Trichoderma reesei* endoglucanase II (EGII, EC 3.2.1.4) was derived from a previous work in our lab with some modifications[27]. The prepro secretion signal peptide, the His epitope tag, and the EGII ORF were kept intact of the original cassette P_{GAL10}-(prepro signal peptide)-His-EGII-docS-T_{ADH1}. The dockerin module docS was replaced with the a-agglutinin mating adhesion receptor to allow

direct anchoring of EGII on the cell surface. The promoter and terminator were changed to P_{TEF1} and T_{PGK1}, respectively, to permit constitutive expression. The EGII cassette was first introduced into the CAD strain on the multi-copy plasmid pRS425. Stained with a monoclonal anti-His antibody and analyzed by flow cytometry, the EGII enzyme was confirmed to be successfully displayed on the yeast cell surface (**Fig. 4.2**). The EGII expression cassette was then integrated into the *leu2* site in the CAD strain to create the CAD-EGII strain.

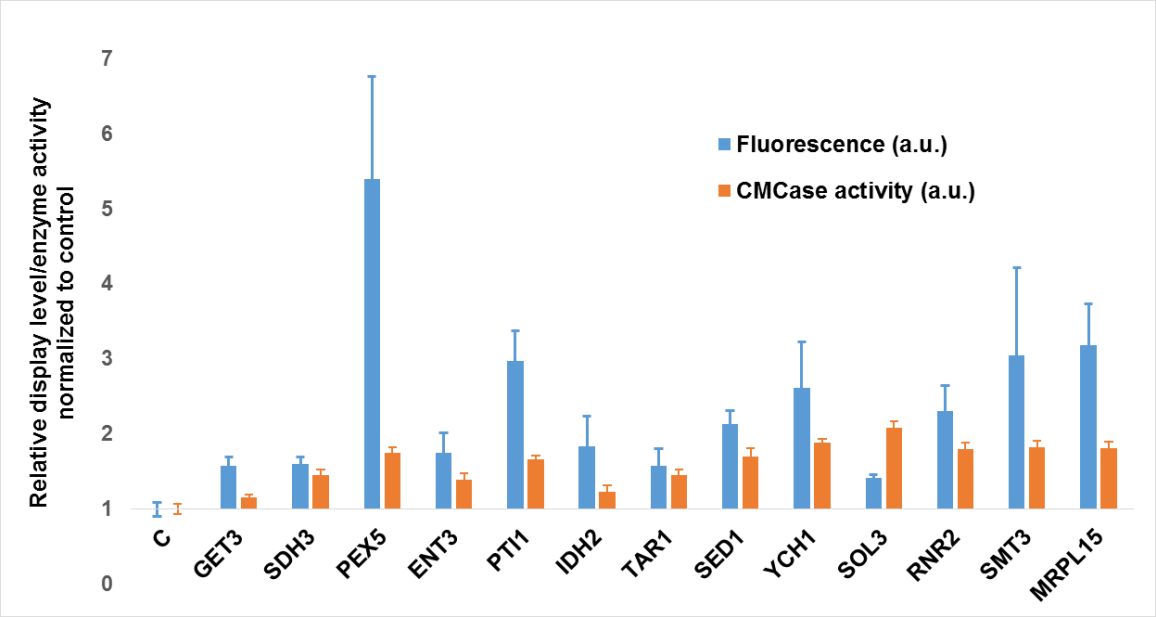

Figure 4.3 Yeast mutant strains identified from the RAGE2.0 library with increased EGII display levels. All the data were normalized relative to the CAD strain containing an empty pRS416 plasmid as the negative control. Error bars indicate standards deviation of biological triplicates.

The RAGE2.0 library constructed on the pRS416 plasmid was transformed into the CAD-EGII strain to obtain $>10^6$ independent clones. After immunostaining, the cells were subjected to two rounds of Fluorescence-Activated Cell Sorting (FACS), and the top 1% brightest cells were collected in each round. The plasmids from 93 mutant strains with the highest fluorescence were isolated and retransformed into the original CAD-EGII strain. After retransformation, flow

cytometry results showed that there were 29 plasmids conferring increased fluorescence compared with the control strain. The insets of the 29 plasmids were identified by DNA sequencing and BLAST search as 13 independent clones, containing 7 overexpression and 6 knockdown cassettes (**Table 4.2** and **Fig. 4.3**). The fact that some cDNA cassettes appeared more than once in the selected collection indicated effective enrichment by FACS (**Table 4.2**). The improved displaying of EGII were also confirmed by enhanced hydrolysis of carboxymethyl cellulose (CMC) observed with the mutant strains (**Fig. 4.3**). Together, these results demonstrated that the RAGE2.0 library was effective in identifying genetic modifications to improve surface display of a recombinant protein.

Table 4.2 Selected genetic targets that improved the surface display level of EGII

Overexpression target		Knockdown target	
Gene name	Frequency	Gene name	Frequency
GET3	3/15	SED1	1/14
SDH3	5/15	YCH1	4/14
PEX5	1/15	SOL3	3/14
ENT3	4/15	RNR2	3/14
PTI1	1/15	SMT3	2/14
IDH2	1/15	MRPL15	1/14
TAR1	1/15		

4.2.2.2 Glycerol utilization

Because of the tremendous growth in biodiesel utilization worldwide, glycerol, a by-product in biodiesel production, is potentially a cheap feedstock for microbial fermentation. Efforts on transformation of glycerol into value-added products include the use of *E. coli* and *S. cerevisiae* to produce ethanol[37, 38], triacylglycerol (TAG)[39] and fatty acid ethyl esters (FAEEs)[40]. Although

S. cerevisiae is able to grow on glycerol as a sole carbon source, the consumption rate and growth rate are quite slow compared to that of glucose[38-40]. Thus, it is needed to improve glycerol utilization in yeast for any efficient fermentation with glycerol as the substrate.

We constructed a RAGE2.0 library in the CAD strain containing $>10^6$ independent clones. Serial transfer was performed along with the control strain in synthetic SC-U medium with 3% glycerol (v/v) as the sole carbon source. When the OD_{600} value reached 1, yeast culture was inoculated into fresh medium using a 1% or 0.1% inoculum (**Table 4.3**). The 0.1% inoculum was used in the last two rounds of subculturing as a more stringent enrichment procedure to minimize false positive[41]. While both the RAGE2.0 library strain and the control strain exhibited accelerated growth during enrichment, the growth rate of the RAGE2.0 library is much faster than the control strain (**Table 4.3**).

Table 4.3 Serial transfer to enrich yeast strains with improved glycerol utilization

	RAGE2.0 library		Control	
	Inoculum (%)	Time to reach OD=1(day)	Inoculum (%)	Time to reach OD=1(day)
Round 1	1	2.5	1	3.5
Round 2	1	2	1	3
Round 3	1	2	1	2.5
Round 4	0.1	2.5	1	2.5
Round 5	0.1	1.5		

The enriched library was streaked onto a solid SC-U glycerol plate. The plasmids from 20 randomly picked colonies were isolated, and the inserts of these plasmids were analyzed by DNA sequencing and BLAST search. There were 4 independent knockdown cassettes identified with some cassettes appearing more than once (**Table 4.4**). After retransformation, all 4

cassettes were confirmed to confer improved glycerol utilization compared to the control strain (**Fig. 4.4**).

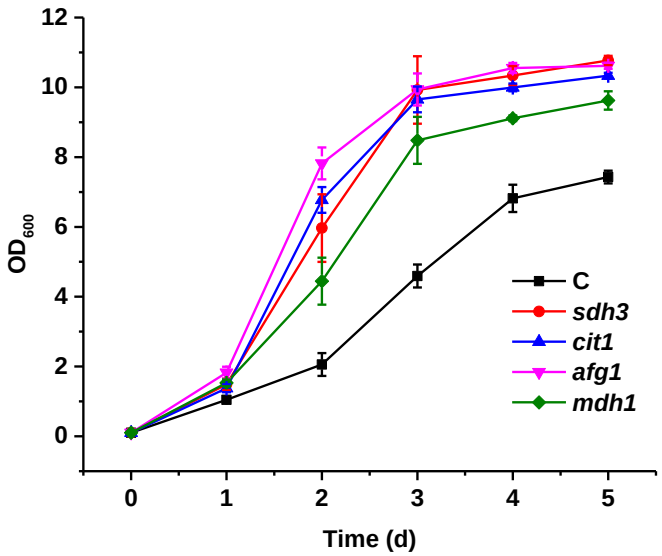

Figure 4.4 Yeast mutant strains identified from the RAGE2.0 library with improved glycerol utilization. Cells were cultured in synthetic dropout medium SC-U with 3% (v/v) glycerol as the sole carbon source. The initial OD600 was 0.1, and the cell density was monitored at the indicated time intervals. The CAD strain containing an empty pRS426 plasmid was used as the control strain. Error bars indicate standard deviations of biological triplicates.

Table 4.4 Selected genetic targets that improved glycerol utilization

Knockdown target	
Gene name	Frequency
SDH3	2/20
CIT1	6/20
AFG1	1/15
MDH1	1/15

4.2.2.3. Isobutanol production

Although ethanol is the most widely used biofuel, advanced biofuels have attracted a growing interest due to their superior fuel properties[42, 43]. In particular, butanol (mainly 1-butanol and isobutanol) is considered a substantially better gasoline substitute than ethanol[44], as it is more hydrophobic and less corrosive, of higher energy density[45], and can be blended in gasoline at any ratio[45]. Specifically, isobutanol can be produced by the Ehrlich pathway in *S. cerevisiae* as a degradation product derived from valine[46]. Intensive engineering of the yeast metabolism has been performed to improve the isobutanol production through rational design[43], yet little efforts have been taken for genome-wide screen for non-intuitive genetic targets.

As isobutanol is derived from valine biosynthesis, construction of valine-overproducing strains seems a feasible route to boost isobutanol production. Resistance towards a toxic analog of an amino acid has been linked to the overproduction of the target amino acid. In particular, valine-overproducing *Corynebacteria* and *E. coli* strains has been evolved when isolating norvaline (NVA) tolerant mutants. The underlying hypothesis is that the toxicity results from incorporation of the analog during protein synthesis, and a cell must overproduce the natural amino acid to outcompete the analog to survive. Although different resistance mechanisms may also arise, such as improved analog efflux, this method has been successfully applied in *E. coli* to isolate mutant strains with improved accumulation of leucine and valine, as well as their derivative fusel alcohols, 3-methyl-1-butanol and isobutanol.

Table 4.5 Serial transfer to enrich yeast strains with improved NVA resistance.

	NVA concentration (g/L)	Time to reach OD_{600}=2.5 (day)	
		RAGE2.0 library	Control
Round 1	4	1.5	2.5
Round 2	4	1	1.5
Round 3	4	1	1.5
Round 4	5	1	1.5
Round 5	5	1	1.5
Round 6	6	1	1.5
Round 7	6	1	1

Here we screened for isobutanol-overproducing *S. cerevisiae* strains from the RAGE2.0 library by selecting NVA-resistant mutants. The same library constructed for glycerol utilization in the CAD strain was used for serial transfer in synthetic dropout medium containing increasing concentrations of NVA from 4 g/L to 6 g/L in seven rounds (**Table 4.5**). For each round, 1% inoculum from the previous round was used to inoculate fresh medium when OD_{600} exceeded 2.5. The growth rates of the RAGE2.0 library and the control strain both increased over time, indicating adaptation of *S. cerevisiae* cells to the NVA toxicity (**Table 4.5**). Nonetheless, the RAGE2.0 library strains grew remarkably faster than the control strain (**Table 4.5**). After subculturing, the enriched library was streaked on SC-U plates containing 6g/L NVA. Twenty colonies were randomly picked and tested for isobutanol production. Only two mutant strains exhibited improved isobutanol production, and the plasmids isolated contained the same anti-sense fragment (2140-2931 bp) of the *BUL1* gene (the ORF length is 2931 bp). After retransformation, this plasmid still conferred a higher isobutanol titer (53.0±1.2 mg/L) compared to the control strain (23.8±2.4 mg/L) (**Fig. 4.5**).

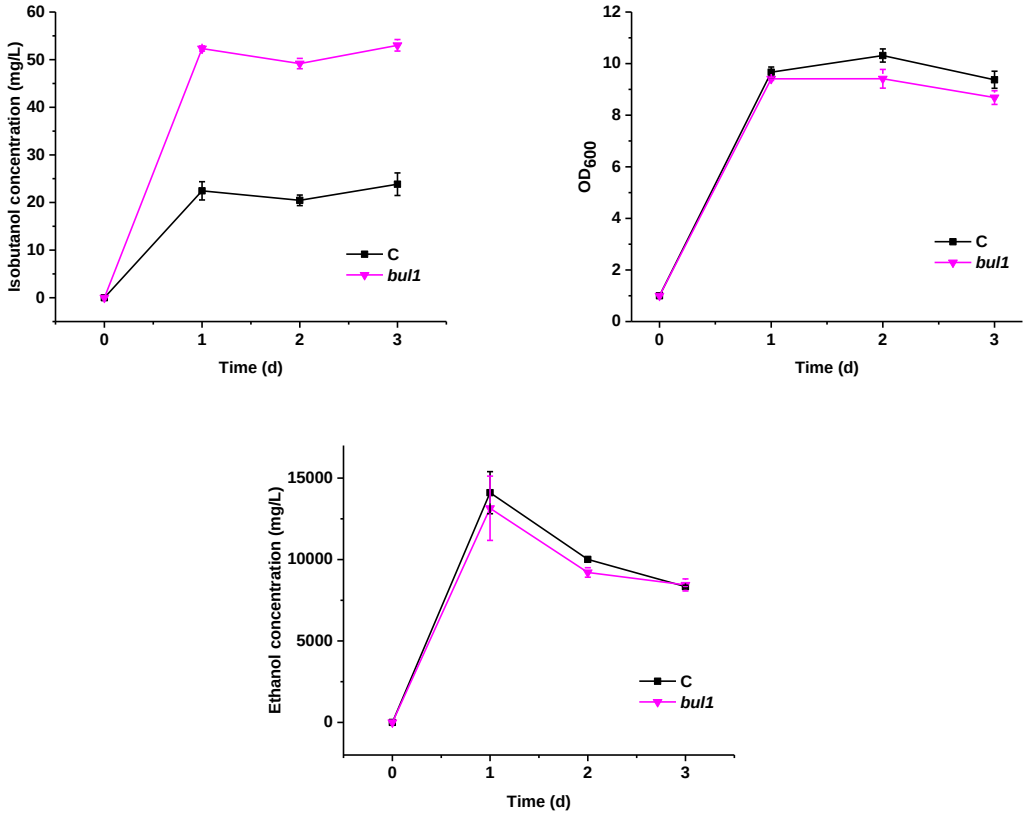

Figure 4.5 Yeast mutant strains identified from the RAGE2.0 library with improved isobutanol production. Cells were cultured under oxygen limited condition in synthetic dropout medium SC-U with 2% glucose. The initial OD600 was 1, and the cell density, ethanol production and isobutanol production were monitored at the indicated time intervals. The CAD strain containing an empty pRS426 plasmid was used as the control strain. Error bars indicate standard deviations of biological triplicates.

4.3. Conclusions and Future Prospects

In this chapter, we devised a strategy to construct a comprehensive library of *S. cerevisiae*. Whereas the cloning of the genome-wide cDNA library under the control of a promoter is a standard method to construct an overexpression library, we made two key modifications, including insertion of the cDNA molecules in a reversed direction and introduction of an RNAi

pathway to successfully adapt the same procedure to construct a knockdown library. Therefore, all the high-throughput strategies used previously to screen the overexpression libraries can be readily applied to the RAGE2.0 library, which can generate more comprehensive knowledge about the genotype-phenotype relationship by including the reduction-of-function mutations. In its current form, the RAGE2.0 library derives from the cDNA library, and thus does not require genome sequence or annotation to create. This feature makes RAGE2.0 a simple and inexpensive solution for genetic screen, especially for less well-studied industrial strains and natural isolates, compared with the traditional genome-wide libraries that are created with the help of many specific DNA primers. But it also suggests that some inherent drawbacks of the cDNA library may limit the use of RAGE2.0. For example, even though the size fractionation and normalization steps can help to enrich those rare and/or long transcripts, some genes may be inevitably missing from the library. In the future, microarray analysis and/or next-generation sequencing (NGS) may be performed to investigate the coverage of the RAGE2.0 library, so that missing cassettes can be manually added through the traditional cloning method.

Coupled with high-throughput screening, the RAGE2.0 library successfully identified genetic modifications to improve a series of phenotypes that are important for microbial fermentation. Although retransformation have showed that the desirable traits were due to the selected cassettes rather than adaptive mutations in the genome, further investigation is needed to confirm that these cassettes mediate the improvement through expected modifications, rather than "off-target" effects. For knockdown targets, in particular, real time RT-PCR will help to estimate the gene repression efficiency by the selected anti-sense constructs, and gene deletion should also be performed to inspect how the target phenotypes are affected. It is also interesting to examine

whether combinatorial optimization among the selected modifications can result in further improvement.

For glycerol utilization and isobutanol production, only knockdown targets were identified. While there is a possibility that these are indeed the genetic modifications conferring the most remarkable improvements, it is more likely that the absence of the overexpression targets are due to the inherent limitation of the growth-based enrichment method. Whereas a knockdown cassette triggers gene repression via transcription of the anti-sense mRNA, an overexpression cassette can only impact the phenotype by its encoded protein, whose expression may exert more burdens on the cell than that of an anti-sense mRNA. Therefore, the growth advantage of an overexpression cassette may be compromised by its extra burden of protein expression. During prolonged serial transfer process, those knockdown mutants may outcompete the overexpression mutants, and hence take over the whole population. This could be one possible reason why no overexpression modifications were isolated to improve glycerol utilization and isobutanol production. Such problems may be solved via the use of a colony size-based selection strategy.

The mechanisms on how some selected cassettes affect the targeted phenotypes can be speculated from their known functions. For example, a mutant strain with an improved EGII display level harbored a knockdown cassette of the *SED1* gene. The *SED1* gene encodes a stress-induced structural cell wall protein Sed1p, which is anchored on the cell surface through a glycosylphosphatidylinositol (GPI) domain[47]. It has been shown that the use of the *SED1* promoter and the Sed1p anchoring domain enabled highly efficient display of heterologous proteins on the cell surface of *S. cerevisiae*[29], suggesting that the expression level of the Sed1p protein is quite high. Hence, it is possible that decreased expression of the Sed1p protein can improve the capacity of the cell surface to accommodate other cell wall proteins, including the

EGII-agglutinin fusion. In addition, three knockdown cassettes conferring improved glycerol utilization all targeted tricarboxylic acid (TCA)-related genes, *SDH3*, *CIT1* and *MDH1*, and the reduced TCA activity has been linked to fast glycerol metabolism in *E. coli*[48]. Moreover, knockdown of the *BUL1* gene may lead to improved isobutanol production through mistargeting of the general amino acid permease Gap1p, whereas reduced expression of the Gap1p was found to increase isobutanol titer[49]. Further mechanistic studies will be helpful to validate these speculations, as well as to understand other genetic modifications that lack intuitive functional explanations.

4.4. Materials and Methods

4.4.1. Strains, media, and cultivation conditions

S. cerevisiae strain CEN.PK2-1c (*MAT**a** ura3-52 trp1-289 leu2-3,112 his3Δ1 MAL2-8C SUC2*) was purchased from EUROSCARF (Frankfurt, Germany). Zymo 5α Z-competent *E. coli* (Zymo Research, Irvine, CA) and NEB 10β Electrocompetent *E. coli* (New England Biolabs, Ipswich, MA) were used for plasmid amplification and library construction, respectively. *S. cerevisiae* strains were cultivated in either synthetic dropout medium (0.17% Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate and 0.083% amino acid drop out mix, 0.01% adenine hemisulfate and 2% glucose) or YPAD medium (1% yeast extract, 2% peptone, 0.01% adenine hemisulfate and 2% glucose). For glycerol utilization, 3% (v/v) glycerol was used instead of 2% glucose in the synthetic dropout medium. The 50 g/L filtered norvaline (NVA) solution stock was added into the synthetic dropout medium to select for NVA resistant strains. Before mixing, both the medium and NVA stock were adjusted to pH = 5.6. *S. cerevisiae* strains were cultured at 30 °C and with 250 rpm agitation in baffled shake-flasks for aerobic growth, and at 30 °C and 100 rpm in un-baffled shake-flasks for oxygen limited fermentation. *E. coli* strains were cultured at 37 °C and 250 rpm in Luria broth (LB) medium (Fisher Scientific, Pittsburgh, PA) with the supplement of 100 μg/mL ampicillin. All chemicals were purchased through Sigma-Aldrich or Fisher Scientific.

4.4.2. DNA manipulation

Plasmid cloning was done by Gibson Assembly Cloning Kit from New England Biolabs following the manufacturer's instructions or by the DNA assembler method[50]. For DNA

manipulations, yeast plasmids were isolated using a Zymoprep II yeast plasmid isolation kit (Zymo Research, Irvine, CA) and transferred into *E. coli* for amplification. QIAprep Spin Plasmid Mini-prep Kits (Qiagen, Valencia, CA) were employed to prepare plasmid DNA from *E. coli*. All enzymes used for recombinant DNA cloning were from New England Biolabs unless otherwise noted. The products of PCR, digestion and ligation reactions were purified by QIAquick PCR Purification and Gel Extraction Kits (Qiagen, Valencia, CA).

4.4.3. Construction of the RAGE2.0 library

The *ccdB* gene was cloned from the plasmid pLRX-ccdB[23] into the P_{TEF1}-T_{PGK1} cassette constructed in the **Chapter 3**[13]. The adapter sequences (indicated as **bold** font) and the *Pme*I sites (indicated as *italic* font) were incorporated by the primers PR608-PR611 (**Fig. 4.1** and **Table 4.6**). The expression cassettes were then cloned into the pRS416 or pRS426 plasmid, and the resultant constructs were digested by the *Pme*I enzyme to serve as the backbone vectors for the RAGE2.0 library.

Table 4.6 Primers used for the RAGE2.0 library construction

Name	Primer sequences (5'→3')	Cassettes (**Fig. 4.1**)
PR608	agttttaattacaaa **aagcagtggtatcaa** *gtttaaac* caggaagggatggctgagg	P_{TEF1p}-ccdB-T_{PGK1}-F
PR609	tcaattcaattcaat **cggggtacgatgaga** *gtttaaac* gagctctagagatatcgtcg	
PR610	agttttaattacaaa **cggggtacgatgaga** *gtttaaac* caggaagggatggctgagg	P_{TEF1p}-ccdB-T_{PGK1}-R
PR611	tcaattcaattcaat **aagcagtggtatcaa** *gtttaaac* gagctctagagatatcgtcg	

The normalized, full-length cDNA library was prepared as follows. The total RNAs from the CEN.PK2-1c strain were isolated using the RNeasy mini kit (QIAGEN, Valencia, CA). A cDNA library was synthesized using the In-Fusion SMARTer Directional cDNA Library

Construction Kit (Clontech Laboratories, Mountain View, CA) with some modifications. In particular, the double-stranded cDNA library was treated using the Trimmer-2 cDNA normalization kit (Evrogen, Moscow, Russia) according to the manufacturer's instructions. Small cDNA fragments were then removed by the size fractionation step in the SMARTer kit. The treated cDNA library was cloned into the linearized backbone vectors using the Gibson assembly cloning kit (New England Biolabs), and the enzymatic mix was electroporated into the NEB 10β *E. coli* strain and amplified on LB+Amp plates. More than 10^6 independent *E. coli* colonies were obtained for both the overexpression and knockdown libraries. The plasmid minipreped from the *E. coli* library was used to transform the CAD or CAD-EGII strains by the standard LiAc/ssDNA/PEG protocol[51] with an optimized condition, where 50 μg plasmid DNA was used to transform 40 OD_{600} unit competent yeast cells by heat shock at 42 °C for 1 hr. More than 10^6 independent yeast clones were obtained for every library using the optimized protocol. The yeast transformants were amplified on SC-U plates for four days before being harvested for screen.

4.4.4. Screen of EGII-displaying mutants

The CAD-EGII strain was constructed by integrating the EGII-displaying cassette into the *leu2* site of the yeast genome with the help of the pRS405 plasmid. Harvested from the SC-U plates, the RAGE2.0 library created in the CAD-EGII strain was used for subsequent screen and analysis. The CAD-EGII strain contained an empty pRS416 plasmid was used as the control strain. The immunostaining assay and flow cytometry analysis were performed as described previously[27]. The PE fluorescence was analyzed with a LSR II Flow Cytometer (BD Biosciences, San Jose, CA). FACS experiments were performed on a BD FACS Aria III cell

sorting system (BD Biosciences, San Jose, CA). In the first round of sorting, 30,000 cells representing the top 1% brightest fluorescence were collected and grown for two days in the SC-U medium. Then the second round of sorting collected 93 individual yeast cells with the top 1% brightest fluorescence into a 96-well microplate. After retransformation in the CAD-EGII strain with a fresh background, 29 plasmids that still conferred an improved PE fluorescence were sent for DNA sequencing analysis. The retransformed yeast mutants were further analyzed by the CMCase assay. Briefly, 5 OD_{600} unit yeast cells from overnight culture in the SC-U medium were washed twice with ddH_2O and resuspend in 1% (w/v) CMC solution (0.1 M sodium acetate, pH=5). After incubation at 30 °C for 16 hr with shaking (250 rpm), the supernatant was used for reducing sugar analysis by a modified DNS method[52]. The released reducing sugar amount from the CMC hydrolysis by EGII was used to quantify the enzyme activity.

4.4.5. Screen of glycerol-utilizing mutants

Harvested from the SC-U plates, the RAGE2.0 library created in the CAD strain was used for enrichment in the synthetic SC-U 3% glycerol medium. The CAD strain contained an empty pRS426 plasmid was used as the control strain. Serial subculturing was performed according to **Table 4.3**, where 1% or 0.1% inoculum was transferred into the fresh SC-U glycerol medium when the cell density exceeded OD_{600}=1 in a previous round. The initial OD_{600} was 0.01 for both the RAGE2.0 library and the control strain. After enrichment, single colonies were obtained by streaking the library onto a SC-U glycerol plate. 20 randomly picked colonies were examined for growth on glycerol in liquid medium. After retransformation in the CAD strain with a fresh background, 20 plasmids conferring improved glycerol utilization than the control strain were sent for DNA sequencing analysis. The retransformed yeast mutants were compared

with the control strain for glycerol utilization via an aerobic growth assay. Overnight culture in the SC-U 2% glucose medium was used to inoculate 3 ml SC-U 3% glycerol medium in 14 ml falcon tubes. The initial OD600 was set to 0.1, and the cell densities were monitored at 24 hr interval under the aerobic condition.

4.4.6. Screen of isobutanol-producing mutants

Harvested from the SC-U plates, the RAGE2.0 library created in the CAD strain was used for enrichment in the synthetic SC-U medium containing increasing concentrations of NVA (**Table 4.5**). The CAD strain contained an empty pRS426 plasmid was used as the control strain. Serial subculturing was performed according to **Table 4.5**, where 1% inoculum was transferred into fresh selective medium when the OD_{600} exceeded 2.5 in a previous round. The initial OD_{600} was 0.01 for both the RAGE2.0 library and the control strain. After the enrichment, single colonies were obtained by streaking the library onto a SC-U plate containing 6 g/L NVA. 20 randomly picked colonies were examined for isobutanol production in 14 mL sealed falcon tubes with 3 mL SC-U medium without NVA. After retransformation in the CAD strain with a fresh background, 2 plasmids conferring improved isobutanol production than the control strain were sent for DNA sequencing analysis. The retransformed yeast mutants were compared with the control strain for isobutanol production. Overnight culture in the SC-U 2% glucose medium was used to inoculate 10 mL SC-U 2% glucose medium in 125 mL un-baffled flasks. The agitation was set to 100 rpm for oxygen limited fermentation. The initial OD600 was 1, and 1 mL culture was taken at 24 hr intervals to measure the cell density, ethanol titer and isobutanol titer. The alcohol concentrations were determined using the GC-MS program described in **Chapter 2**[53].

4.5. References

1. Lynch, M.D., Warnecke, T. & Gill, R.T. SCALEs: multiscale analysis of library enrichment. *Nat Methods* **4**, 87-93 (2007).
2. Ho, C.H. et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol* **27**, 369-377 (2009).
3. Badarinarayana, V. et al. Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat Biotechnol* **19**, 1060-1065 (2001).
4. Alexeyev, M.F. & Shokolenko, I.N. Mini-Tn10 transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of gram-negative bacteria. *Gene* **160**, 59-62 (1995).
5. Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391 (2002).
6. Winzeler, E.A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).
7. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).
8. Meng, J. et al. A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing Escherichia coli essential genes. *FEMS Microbiol Lett* **329**, 45-53 (2012).
9. Wang, B. & Kuramitsu, H.K. Inducible antisense RNA expression in the characterization of gene functions in Streptococcus mutans. *Infect Immun* **73**, 3568-3576 (2005).
10. Forsyth, R.A. et al. A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol Microbiol* **43**, 1387-1400 (2002).

11. Boutros, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nat Rev Genet* **9**, 554-566 (2008).

12. Echeverri, C.J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7**, 373-384 (2006).

13. Si, T., Luo, Y., Bao, Z. & Zhao, H. RNAi-assisted genome evolution in S*accharomyces cerevisiae* for complex phenotype engineering. *ACS Synth Biol*, in press (2014).

14. Breslow, D.K. et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* **5**, 711-718 (2008).

15. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**, 856-862 (2010).

16. Gubler, U. & Hoffman, B.J. A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-269 (1983).

17. Wellenreuther, R., Schupp, I., Poustka, A. & Wiemann, S. SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC Genomics* **5**, 36 (2004).

18. Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. & Siebert, P.D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892-897 (2001).

19. Bogdanova, E.A., Shagin, D.A. & Lukyanov, S.A. Normalization of full-length enriched cDNA. *Mol Biosyst* **4**, 205-212 (2008).

20. Carninci, P. et al. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res* **10**, 1617-1630 (2000).

21. Zhulidov, P.A. et al. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res* **32**, e37 (2004).

22. Shagin, D.A. et al. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Res* **12**, 1935-1942 (2002).

23. Liang, J., Chao, R., Abil, Z., Bao, Z. & Zhao, H. FairyTALE: A High-Throughput TAL Effector Synthesis Platform. *ACS Synth Biol* (2013).

24. Lynd, L.R., van Zyl, W.H., McBride, J.E. & Laser, M. Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol* **16**, 577-583 (2005).

25. Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**, 506-577 (2002).

26. Fujita, Y., Ito, J., Ueda, M., Fukuda, H. & Kondo, A. Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. *Appl Environ Microbiol* **70**, 1207-1212 (2004).

27. Wen, F., Sun, J. & Zhao, H. Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. *Appl Environ Microbiol* **76**, 1251-1260 (2010).

28. Sun, J., Wen, F., Si, T., Xu, J.H. & Zhao, H. Direct conversion of xylan to ethanol by recombinant *Saccharomyces cerevisiae* strains displaying an engineered minihemicellulosome. *Appl Environ Microbiol* **78**, 3837-3845 (2012).

29. Inokuma, K., Hasunuma, T. & Kondo, A. Efficient yeast cell-surface display of exo- and endo-cellulase using the SED1 anchoring region and its original promoter. *Biotechnol Biofuels* **7**, 8 (2014).

30. Kjeldsen, T. et al. Engineering-enhanced protein secretory expression in yeast with application to insulin. *J Biol Chem* **277**, 18245-18248 (2002).

31. Shusta, E.V., Holler, P.D., Kieke, M.C., Kranz, D.M. & Wittrup, K.D. Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat Biotechnol* **18**, 754-759 (2000).

32. Shusta, E.V., Raines, R.T., Pluckthun, A. & Wittrup, K.D. Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat Biotechnol* **16**, 773-777 (1998).

33. Robinson, A.S., Hines, V. & Wittrup, K.D. Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Biotechnology* **12**, 381-384 (1994).

34. Wentz, A.E. & Shusta, E.V. A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl Environ Microbiol* **73**, 1189-1198 (2007).

35. Ramjee, M.K., Petithory, J.R., McElver, J., Weber, S.C. & Kirsch, J.F. A novel yeast expression/secretion system for the recombinant plant thiol endoprotease propapain. *Protein Eng* **9**, 1055-1061 (1996).

36. Harmsen, M.M., Bruyne, M.I., Raue, H.A. & Maat, J. Overexpression of binding protein and disruption of the PMR1 gene synergistically stimulate secretion of bovine prochymosin but not plant thaumatin in yeast. *Appl Microbiol Biotechnol* **46**, 365-370 (1996).

37. Shams Yazdani, S. & Gonzalez, R. Engineering *Escherichia coli* for the efficient conversion of glycerol to ethanol and co-products. *Metab Eng* **10**, 340-351 (2008).

38. Yu, K.O., Kim, S.W. & Han, S.O. Engineering of glycerol utilization pathway for ethanol production by *Saccharomyces cerevisiae*. *Bioresour Technol* **101**, 4157-4161 (2010).

39. Yu, K.O. et al. Development of a *Saccharomyces cerevisiae* strain for increasing the accumulation of triacylglycerol as a microbial oil feedstock for biodiesel production using glycerol as a substrate. *Biotechnol Bioeng* **110**, 343-347 (2013).

40. Yu, K.O., Jung, J., Kim, S.W., Park, C.H. & Han, S.O. Synthesis of FAEEs from glycerol in engineered *Saccharomyces cerevisiae* using endogenously produced ethanol by heterologous expression of an unspecific bacterial acyltransferase. *Biotechnol Bioeng* **109**, 110-115 (2012).

41. Jin, Y.S., Alper, H., Yang, Y.T. & Stephanopoulos, G. Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach. *Appl Environ Microbiol* **71**, 8249-8256 (2005).

42. Peralta-Yahya, P.P., Zhang, F.Z., del Cardayre, S.B. & Keasling, J.D. Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320-328 (2012).

43. Buijs, N.A., Siewers, V. & Nielsen, J. Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Curr Opin Chem Biol* **17**, 480-488 (2013).

44. Jin, C., Yao, M.F., Liu, H.F., Lee, C.F.F. & Ji, J. Progress in the production and application of n-butanol as a biofuel. *Renew Sust Energ Rev* **15**, 4080-4106 (2011).

45. Dürre, P. Biobutanol: An attractive biofuel. *Biotechnology J* **2**, 1525-1534 (2007).

46. Hazelwood, L.A., Daran, J.M., van Maris, A.J., Pronk, J.T. & Dickinson, J.R. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl Environ Microbiol* **74**, 2259-2266 (2008).

47. Shimoi, H., Kitagaki, H., Ohmori, H., Iimura, Y. & Ito, K. Sed1p is a major cell wall protein of *Saccharomyces cerevisiae* in the stationary phase and is involved in lytic enzyme resistance. *J Bacteriol* **180**, 3381-3387 (1998).

48. Martinez-Gomez, K. et al. New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microb Cell Fact* **11**, 46 (2012).

49. Dundon, C.A., Smith, C., Nahreini, P., Thevelein, J. & Saerens, S. Isobutanol Production Using Yeasts With Modified Transporter Expression. *United States patent* **US 13701247 A2** (2013).

50. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**, e16 (2009).

51. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protocols* **2**, 31-34 (2007).

52. King, B.C., Donnelly, M.K., Bergstrom, G.C., Walker, L.P. & Gibson, D.M. An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. *Biotechnol Bioeng* **102**, 1033-1044 (2009).

53. Si, T., Luo, Y., Xiao, H. & Zhao, H. Utilizing an endogenous pathway for 1-butanol production in *Saccharomyces cerevisiae*. *Metab Eng* **22**, 60-68 (2014).

Chapter 5 RAGE3.0: Automated Genome Engineering in Yeast

5.1. Introduction

Laboratory automation can greatly accelerate fundamental and applied research, by improving productivity and reliability, increasing throughput, and reducing experimental error rates due to human factors[1]. For example, pharmaceutical industry has heavily relied on automation technologies to identify new drug lead compounds by screening small-molecule libraries[1]. For biological system engineering, the promise of automation has been demonstrated by several recent examples[2, 3]. Phage-assisted continuous evolution (PACE) executed 200 rounds of protein evolution in 8 days, during which targeted activities effectively emerged from undetectable levels[2]. Moreover, multiplex automated genome engineering (MAGE) created over 4.3 billion combinatorial variants per day, enabling 5-fold increase in lycopene production in 3 days[3]. By eliminating human intervention, rapid and continuous generation of vast diversity space help to fully harness the power of directed evolution.

In previous chapters, we developed plasmid-borne RNAi libraries for genetic screen in *S. cerevisiae*. For directed genome evolution by RAGE, a separate step is required to integrate the selected cassettes from the plasmid library[4], hence manual intervention is unavoidable. To automate genome engineering by RAGE, it is desirable to combine the screen and integration steps by directly integrating genome-wide modifying cassettes into the yeast genome for library construction. Nonetheless, while the transformation efficiency of an episomal plasmid can easily achieve a complete coverage of the entire genome, the integration efficiency needs to be greatly improved to create a comprehensive library.

It has been shown that introduction of DSBs can greatly enhance integration efficiency by stimulating homologous recombination (HR) rates. In *S. cerevisiae*, the *Delitto Perfetto* system was used to improve targeted genome integration by 4000-fold, where a single DSB was created by an inducible I-SceI endonuclease and subsequently repaired with integrative recombinant oligonucleotides[5]. Though effective, an I-SceI recognition site must be first introduced at the targeted genomic loci through the traditional methods, and this prerequisite limits the general application of this method. Recently, the bacterial type II CRISPR-Cas system has been used in *S. cerevisiae* for efficient genome editing[6]. The Cas9 nuclease and designer CRISPR guide RNAs (gRNAs) created specific genomic DSBs, which increased HR rates by 130-fold with double-strand oligonucleotide donors. More impressively, if the gRNA was maintained on a plasmid and the Cas9 protein was constitutively expressed, the wild-type cells carrying targeted genomic sequences were negatively selected. Therefore, only recombinant strains whose targeted sequences were modified by the donor DNA can survive, resulting in near 100% recombination frequency[6]. However, the mutations introduced by this method are limited to the scale of several nucleotides that are embedded in the donor oligonucleotides. Whether CRISPR-mediated DSBs can increase the integration efficiency of large DNA constructs in *S. cerevisiae* is unclear.

Delta (δ)-integration is an effective method for stable and multi-copy integration in *S. cerevisiae*[7-9]. The δ sequence in the yeast retrotransposon elements Ty1 and Ty2 is the homolog of the mammalian retroviral LTR[10]. Approximately 250 copies of the δ sites are dispersed throughout the *S. cerevisiae* genome[11]. Hence, the δ sites are the ideal targets for high copy number integration, which is desirable for high level expression of a recombinant protein in *S.*

cerevisiae[12-14]. In addition, cocktail δ-integration has been used to introduce different genes simultaneously to construct metabolic pathways, including a three-gene cellulolytic pathway[15] and a five-gene isobutanol pathway[16]. Therefore, it is possible to accommodate multiplex integration in the δ sites, allowing continuous accumulation of beneficial RAGE cassettes.

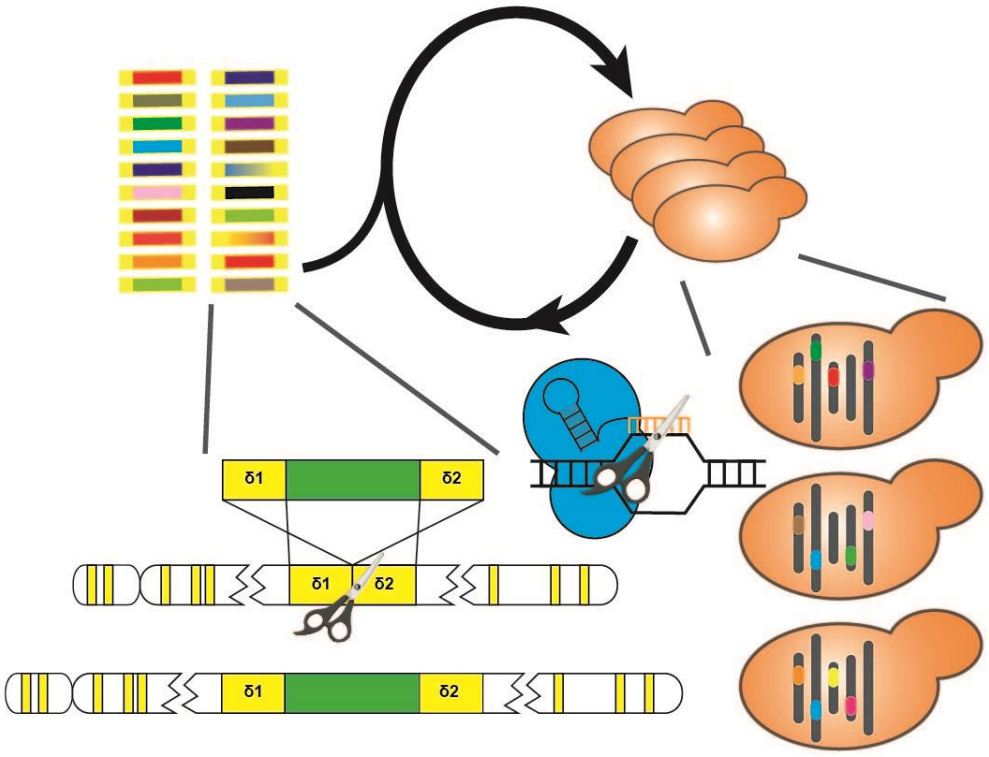

Figure 5.1. Scheme of automated genome engineering in *S. cerevisiae* by RAGE3.0. CRIPSR-assisted δ-integration is used to continuously incorporate genome-wide modifying cassettes in the yeast genome.

In this chapter, we sought to develop an automated genome engineering method in *S. cerevisiae* by combining RNAi, CRISPR and δ-integration technologies. An inducible CRISPR-Cas system was used to introduce specific DSBs in the δ sites. A GFP-based reporter system was then used to optimize the performance of CRISPR-assisted δ-integration. Finally, genome-wide modifying cassettes from the RAGE2.0 library were continuously and iteratively integrated

into the δ sites (**Fig. 5.1**), enabling isolation of fast glycerol-utilizing mutants integrated with multiple RAGE2.0 cassettes.

5.2. Results

5.2.1. Design of RAGE3.0

A δ sequence derived from the *LTR2* element of a yeast transposon Ty1-H3 is used to design the CRISPR targeting site (δ.a) and the homologous arms (**Fig. 5.2**). The δ.a sequence, 5'-GAAGTTCTCCTCAAGGATTTAGG-3', contains a 20 bp target site followed by "AGG" as the required protospacer adjacent motifs (PAM)[17]. The δ.a sequence is flanked by two homologous arms, δ1 and δ2, and insertion of a donor cassette will remove the δ.a sequence to avoid CRISPR recognition after integration (**Fig. 5.2**). The expression cassette of the RAGE2.0 library is used, so that comprehensive genome-wide modifying cassettes can be cloned as the integration donors (**Fig. 5.3**).

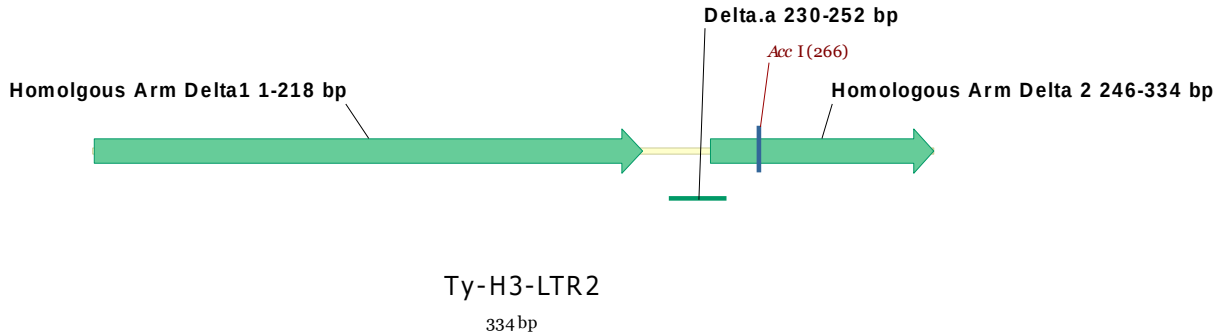

Figure 5.2. Design of a CRISPR target site and homologous arms for δ-integration.

A homology-integrated CRISRP-Cas (HI-CRISPR) system, which was recently developed in our group[18], was exploited to introduce DSBs within the δ sites. In HI-CRISPR, the donor

cassette was maintained on a plasmid to improve HR rate. The Cas9 expression cassette was integrated into the *leu2* locus of the CAD strain to create the CAD-P_{ADH2}-Cas9 strain. The alcohol dehydrogenase II promoter (P_{ADH2}) was used to drive inducible and high level expression of Cas9. A multifunctional donor plasmid was constructed (**Fig. 5.3**), consisting of the targeting CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) cassettes to direct Cas9 to the δ.a sequence, a 2-micron replication origin to ensure high plasmid copy number, a *KanMX* selection marker to maintain the plasmid during induction in rich medium by the antibiotic G418, a *URA3* selection marker for selection and curation of the plasmid, and an integration donor cassette. In addition, the δ.a sequence was also incorporated immediately outside the homologous arms of the donor cassette (**Fig. 5.3**), so that linear integration cassette can be released *in vivo* from the donor plasmid upon digestion by Cas9.

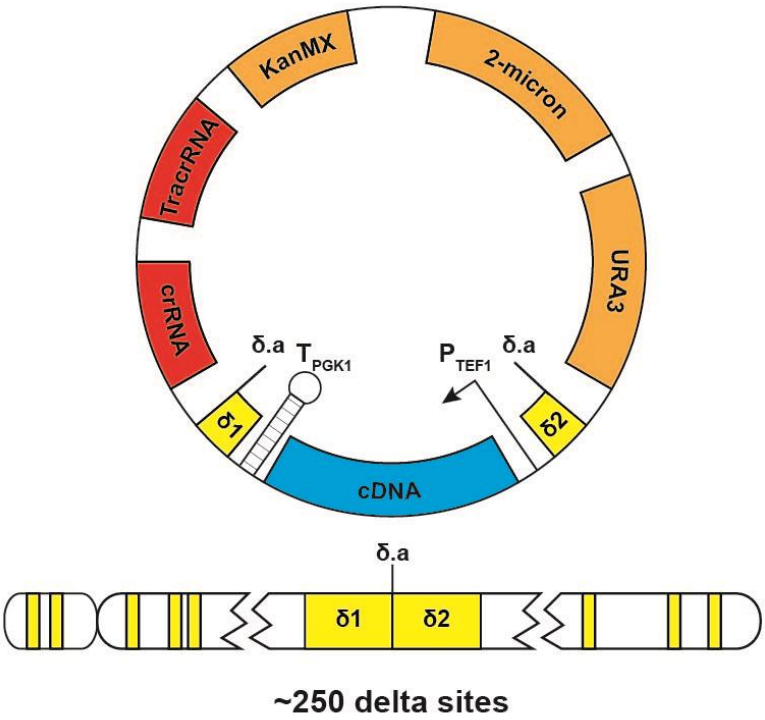

Figure 5.3. Design of the multifunctional integration donor plasmid

A complete cycle of RAGE3.0 library construction comprises three steps: transformation, induction and plasmid curation (**Fig. 5.4**). First, the donor plasmid library harboring the genome-wide modifying cassettes (RAGE2.0) is transformed into the CAD-P_{ADH2}-Cas9 strain. During competent cell preparation and transformant selection, synthetic dropout medium is used to minimize Cas9 expression from the P_{ADH2} promoter. After 3 or 4 days, transformed cells are transferred into the YPA rich medium containing 2% galactose and 1 g/L G418, so that the glucose-repressed P_{ADH2} promoter can be induced to express the Cas9 nuclease at high levels. Following induction, the cells are cultured in the synthetic dropout medium containing 5-fluoroorotic acid (5-FOA) for plasmid curation. The cells grown in the 5-FOA medium are then ready for the next round of RAGE3.0.

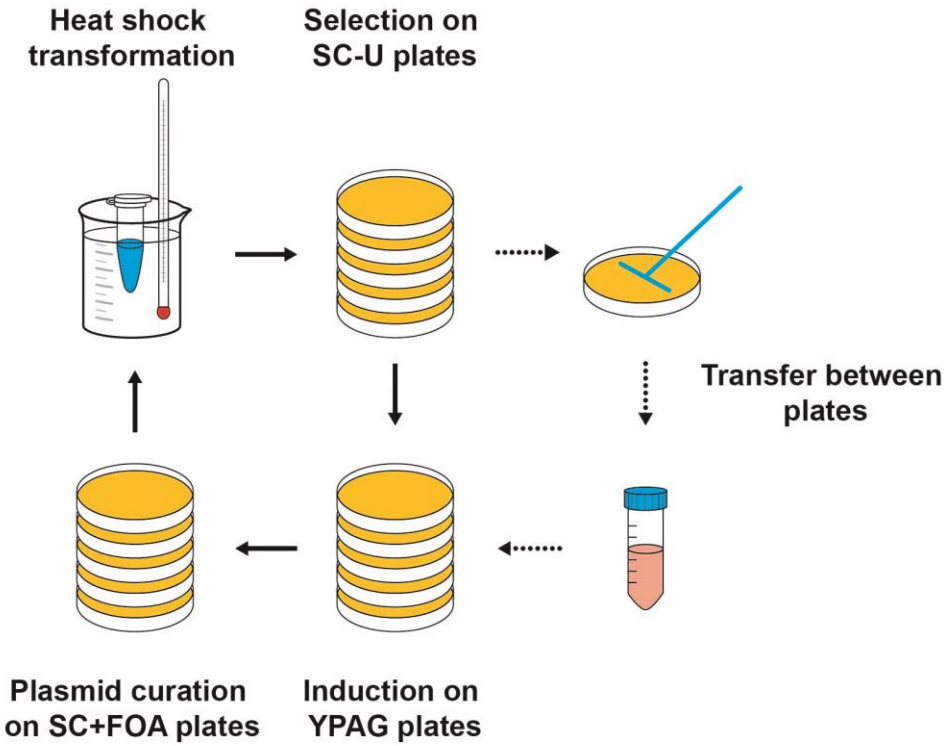

Figure 5.4. Scheme of RAGE3.0 process. Though manipulation is shown with solid medium in petri dishes, it is also possible to perform these steps in liquid medium.

5.2.2. Highly efficient δ-integration of a GFP reporter

To optimize CRIRSPR-assisted δ-integration, a GFP-based reported system was devised. The *gfp* gene was inserted in the P_{TEF1}-T_{PGK1} cassette of the donor plasmid (**Fig. 5.3**), and the resultant GFP-integration donor plasmid was used to perform one cycle of RAGE3.0 (**Fig. 5.4**). After plasmid curation, only the cells integrated with the GFP donor cassette retain GFP fluorescence, hence the percentage of GFP positive population is a fair estimation of the integration efficiency. To ensure high expression level of Cas9, a modified protocol different from that in **Fig. 5.4** was used. In particular, competent cells were prepared using the rich medium YPAD, and the transformants were directly selected in the YPAD medium containing 1 g/L G418. When glucose was depleted, Cas9 expression from P_{ADH2} was automatically induced in the same medium. After plasmid curation in the SC+5-FOA medium, flow cytometry analysis indicated that very high integration efficiency was achieved (>90%) in the CAD-P_{ADH2}-Cas9 strain (**Fig. 5.5**). In comparison, there was no GFP positive population in the CAD strain containing no Cas9 expression cassette, indicating effective curation of the donor plasmid by counter-selection with 5-FOA. If Cas9 expression was driven by a constitutive P_{TEF1} promoter, the GFP positive percentage was remarkably lower than that of the CAD-P_{ADH2}-Cas9 strain, probably due to the weaker promoter strength of P_{TEF1} compared to P_{ADH2}. Together, an efficient Cas9-dependent δ-integration of a GFP donor cassette was successfully demonstrated.

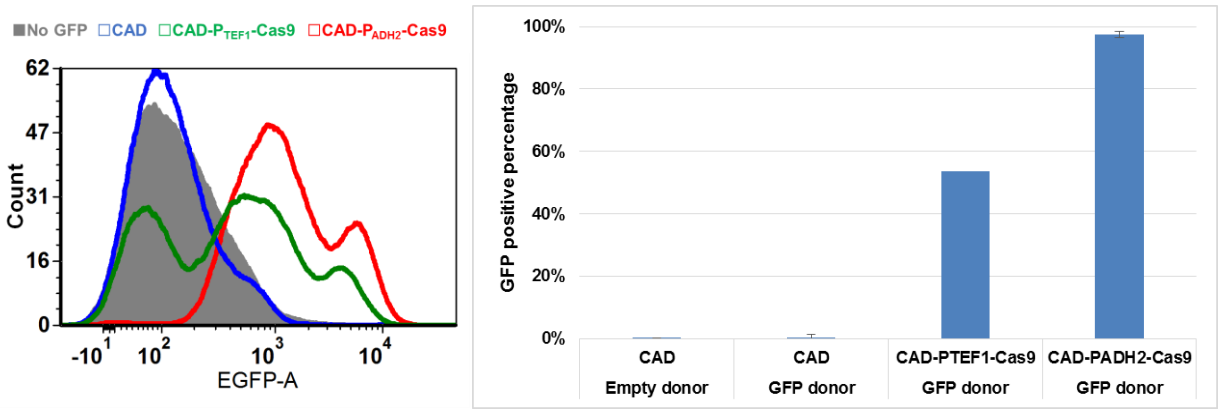

Figure 5.5. Cas9-dependent δ-integration of a GFP donor cassette. Left Panel: flow cytometry histogram. Right Panel: the percentages of GFP positive population. Error bars indicate standard deviation from biological duplicates. No GFP control (grey fill) was the CAD strain transformed with an empty donor plasmid. For all other strains, the donor plasmid containing a GFP expression cassette was used. Blue line: the CAD strain. Green line: the CAD-P_{TEF1}-Cas9 strain. Red line: the CAD-P_{ADH2}-Cas9 strain.

Although high integration efficiency was observed in the CAD-P_{ADH2}-Cas9 strain with the modified protocol, the transformation efficiency was extremely low (<10^2 CFUs/ μg DNA). This is probably due to the fact that the P_{ADH2} promoter is not strictly repressed by glucose in rich medium[19]. It is speculated that under this condition, Cas9 continuously creates DSBs at the δ.a site on both the donor plasmid and the genome, resulting in cellular toxicity that prevents recovery of transformants. Instead, if synthetic dropout medium containing 2% glucose was used to prepare competent cells and to select transformants (**Fig. 5.4**), the obtained transformation efficiency was comparable to the levels routinely achieved by heat shock transformation (10^4 CFUs/ μg DNA). This is consistent with a previous report that gene expression driven by P_{ADH2} is at a minimal level in synthetic glucose medium[19]. After selection in the SC-U medium, cells were transferred into YPA medium containing 1 g/L G418 and different carbon sources for induction, including 2% glucose, 2% ethanol and 2% galactose,

followed by plasmid curation in the SC+5-FOA medium. A minimal level of GFP integration (<2%) was observed if the SC-U 2% glucose medium was used during the induction step (**Fig. 5.6**, grey fill), confirming the relatively stringent repression of P_{ADH2} promoter in the synthetic medium[19]. Among different induction media, the most efficient integration was achieved with the YPAG induction (**Fig. 5.6**, blue line), though the efficiency was a little lower (~70%) compared to that in **Fig. 5**.5. However, as a high transformation efficiency is required to create genome-wide strain libraries, the current protocol represents a good balance between library size and integration efficiency.

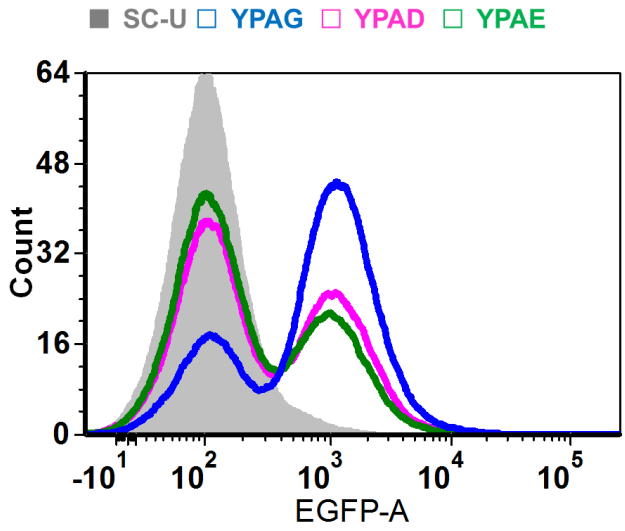

Figure 5.6. Comparison of different induction media on CRISPR-assisted δ-integration.

To further investigate whether iterative cycles of transformation, induction and plasmid curation can enable accumulation of integration cassettes, an additional round of RAGE3.0 was performed with the GFP donor plasmid. Seen from the flow cytometry results, the percentage of GFP population and the collective GFP fluorescence both increased after the second round of

RAGE3.0 process (**Fig. 5.7**), validating the effectiveness of RAGE3.0 to accumulate multiple genetic modifications in the yeast genome.

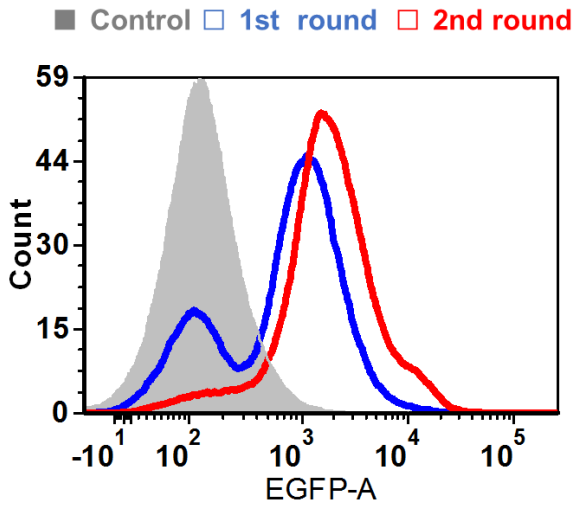

Figure 5.7. Accumulation of GFP integration cassette by RAGE3.0. The control strain is the CAD-P_{ADH2}-Cas9 strain (grey fill).

5.2.3. RAGE3.0 to improve glycerol utilization

Encouraged by the results with the GFP reporter, we employed the RAGE2.0 library for automated genome engineering via CRISPR-assisted δ integration. Three rounds of RAGE3.0 were performed according to **Fig. 5.4**, and the solid medium was used to minimize uneven amplification of the strain library. For each transformation steps, $>10^6$ independent clones were obtained by transforming 100 μg donor library DNA. For each transfer step between plates, 20 OD unit yeast cells (about 10^8 clones) were spread onto twenty 15 mm diameter petri dish to ensure oversampling of the library.

After three rounds of RAGE3.0, 10^7 cells were screened on twenty 15 mm diameter SC-L plates containing 3% glycerol as the sole carbon source. Twenty colonies with remarkably bigger sizes were selected for subsequent analysis. The glycerol utilization capacity was investigated using a growth assay in SC-L 3% glycerol liquid medium. All the selected strains exhibited substantially faster growth rates compared to the CAD-P_{ADH2}-Cas9 strain as the control strain.

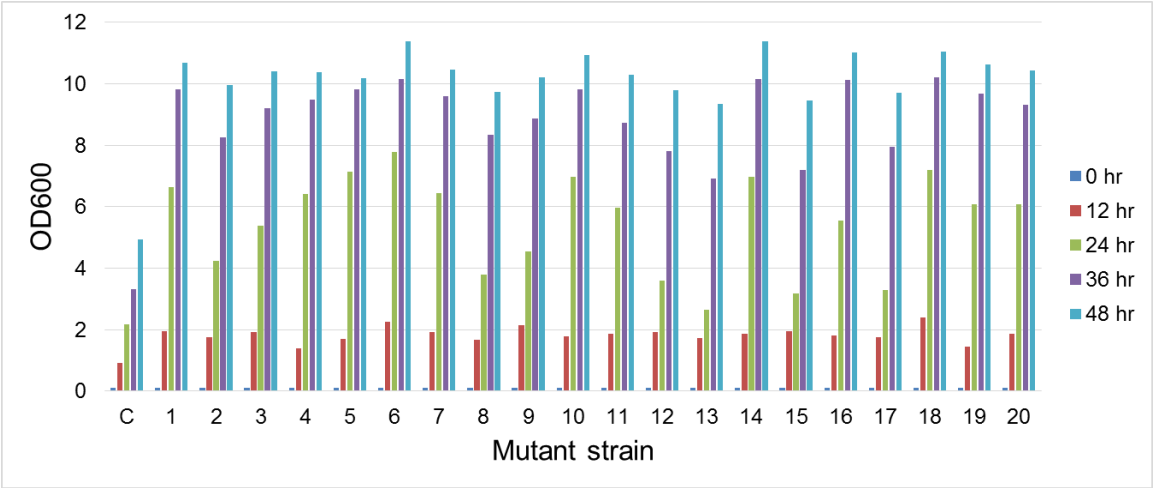

Figure 5.8. Growth of RAGE3.0 mutant strains on glycerol as the sole carbon source. The strains were grown aerobically in 3 mL SC-L 3% (v/v) glycerol medium. The initial OD_{600} is 0.1. The CAD-P_{ADH2}-Cas9 strain was used as the control strain.

For the top 3 mutants with the highest growth rates (G5, G6 and G10), specific primers were used to amplify the integrated cassettes by PCR using genomic DNA as templates. The PCR product was cloned into an *E. coli* plasmid, and 20 individual *E. coli* colonies was picked for each strain to prepare the plasmids for DNA sequencing. The sequencing results indicated that multiple RAGE2.0 cassettes were integrated in each strain (**Table 5.1**). Some common cassettes were identified from different strains, including a knockdown cassette for the *BUL1* gene in the G5 and G10 strains, and a knockdown cassette for the *AFG1* gene in the G5 and G6 strains.

Notably, the same knockdown cassette for the *AFG1* gene was also identified from the RAGE2.0 library (**Fig. 4.3**), suggesting the screen with the RAGE3.0 library for improved glycerol utilization was effective.

Table 5.1. Integration cassettes of RAGE3.0 mutants with improved glycerol utilization

Overexpression		Knockdown	
Plasmid	Inset information	Plasmid	Inset information
G5			
G5_2-F	*CPR3* 1-549/549*	G5_11-R	*AFG1* 1-449/1530
G5_5-F	IV 130506-130188**	G5_12-R	XIV 284075-284374
G5_7-F	XIV 629601-629503	G5_15-R	*BUL1* 2140-2931/2931
G5_9-F	*UBR1* 5453-4982/5853		
G6			
G6_1-F	*RKI1* 1-777/777	G6_14-R	*MOS1* 1-294/294
G6_4-F	*MRPL36* 1-534/534	G6_15-R	XIII 28149-27978
G6_5-F	*PFY1* 1-381/381	G6_17-R	*YSF3* 1-258/258
G6_7-F	*HYP2* 1-474/474	G6_18-R	*AFG1* 1-449/1530
G6_8-F	*ERG6* 829-1152/1152[$]	G6_19-R	*YLR446W* 1-408/1302
G10			
G10_1-F	*RPL10* 1-666/666	G10_11-R	*BUL1* 2140-2931/2931
G10_2-F	*STE4* 435-1272/1272[$]	G10_12-R	*SOD1* 1-465/465
G10_4-F	*RPP1A* 1-321/321	G10_13-R	V 130506-130188
G10_5-F	*FMC1* 1-468/468	G10_18-R	XVI 436924-437120

* For the genes, A-B/X indicates that the inset corresponds to the A-B bp region of the gene's ORF, whose length is X bp. For overexpression cassettes, A indicates the nearest nucleotide to the P_{TEF1} promoter, while for knockdown cassettes, A indicates the nearest nucleotide to the T_{PGK1} terminator.

** If the fragment is not from an ORF, the location is denoted as a reference to the chromosome.

[$] Some overexpression cassettes do not contain complete ORFs.

5.3. Conclusions and Future Prospects

In this chapter, we successfully developed a highly efficient δ-integration method to enable automated genome engineering in *S. cerevisiae*. In a single round of transformation, more than 10^6 independent clones can be obtained, and about 70% cells are modified without any selection marker. Coupled with the comprehensive genome-wide modulating cassettes from the RAGE2.0 library, continuous genome evolution has been achieved. Compared to another automated genome engineering method MAGE in *E. coli*[3], RAGE3.0 achieved substantial improvement in many aspects, including but are limited to recombination efficiency (MAGE can modify 30% cells per single transformation), and the scale of genetic targets (from dozens of predefined genomic loci by MAGE to the genome-wide modifications by RAGE3.0). Together, this is the first example of automated genome engineering in a eukaryote.

There are two search dimensions for genetic modifications conferring an improved trait. First, genotyping to find relevant genes of a given phenotype, mostly by screening genome-wide strain libraries including overexpression[20, 21], knockout[22-24] and knockdown[4, 25] modifications. Second, fine-tuning gene expression of many relevant targets simultaneously to search for the best combinations, mostly via combinatorial modulation on promoters[26], RBSs[3] and intergenic regions[27] of predefined gene sets. Theoretically, some positive genetic modifications can only be identified when these two search efforts are performed at the same time. These modifications may be moderately beneficial, neutral or even detrimental to a given trait when introduced individually, but will confer a remarkably improved phenotype when introduced simultaneously. Different strategies have been used to identify these beneficial genetic combinations. For example, double-mutation strain libraries have been constructed by the synthetic genetic array (SGA) method to systematically investigate genetic interaction network[28-30]. Global

Transcriptional Machinery Engineering (gTME) introduces mutations into key transcription factors to modulate hundreds of genes simultaneously[31, 32]. RAGE3.0 may also serve to identify beneficial combinations of genetic modifications, as multiplex modifying cassettes can be incorporated in a single cell (**Table 5.1**). This multiplex property can be attributed to two factors. First, several donor plasmids can be taken up by a single cell during transformation, and the 2-micron replication origin allows co-existence of multiple plasmids for subsequent integration. Second, the RAGE3.0 cycle can be iteratively performed with minimal human intervention, hence integration cassettes from different rounds can be accumulated.

It should be noted that due to the vast combinatorial space of multiple genome-wide modifications, it is not possible to cover all the combinations even only considering digenic pairs (~6000 yeast genes give rise to ~10^4 single overexpression/knockdown modifications and ~10^8 double mutations, while the library size in each round is about 10^6). All the automated genome engineering strategies will inevitably face this limitation due to finite transformation efficiency, although it has been proposed that through successive cycles of diversity generation, more variants can be cumulatively generated to exceed the actual size of the cell population in a single cycle[3]. The same argument is also valid for RAGE3.0. As RAGE3.0 involves only simple liquid handling steps, it is readily automated by a robotic liquid handling station for continuous genome engineering practice. On the other hand, even the library size at any single round may reflect only a subset of all the theoretically possible variants, beneficial genetic combinations may still be identified. For example, 170,000 interactions were identified from 5.4 million digenic pairs in *S. cerevisiae*[30] (by crossing 1711 queries to 3885 array strains), indicating the prevalence of genetic interactions (3%). Therefore, it may be possible to isolate a few synergistic modifications from a 10^6 library. In the future, it will be helpful to further analyze how multiple

genetic modifications collectively or synergistically confer improved glycerol utilization in the mutant strains isolated in the chapter.

5.4. Methods and Materials

5.4.1 Strains, media, and cultivation conditions

S. cerevisiae strain CEN.PK2-1c (*MAT**a** ura3-52 trp1-289 leu2-3,112 his3Δ1 MAL2-8C SUC2*) was purchased from EUROSCARF (Frankfurt, Germany). The CAD strain containing a reconstituted RNAi pathway was constructed previously[4]. Zymo 5α Z-competent *E. coli* (Zymo Research, Irvine, CA) and NEB 10β Electrocompetent *E. coli* (New England Biolabs, Ipswich, MA) were used for plasmid amplification and library construction, respectively. *S. cerevisiae* strains were cultivated in either synthetic dropout medium (0.17% Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate and 0.083% amino acid drop out mix, 0.01% adenine hemisulfate and 2% glucose) or YPAD medium (1% yeast extract, 2% peptone, 0.01% adenine hemisulfate and 2% glucose). For glycerol utilization, 3% (v/v) glycerol was used instead of 2% glucose in the synthetic dropout medium. For induction, 2% ethanol or 2% galactose was used instead of 2% glucose in the YPA medium. The induction medium was also supplemented with 1 g/L G418 to maintain the donor plasmid. For plasmid curation, 1 g/L 5-fluoroorotic acid (5-FOA) was added into the synthetic dropout medium (pH=4). *S. cerevisiae* strains were cultured at 30 °C and with 250 rpm agitation in baffled shake-flasks for aerobic growth, and at 30 °C and 100 rpm in un-baffled shake-flasks for oxygen limited fermentation. *E. coli* strains were cultured at 37 °C and 250 rpm in Luria broth (LB) medium (Fisher Scientific, Pittsburgh, PA) with the supplement of 100 μg/mL ampicillin. All chemicals were purchased through Sigma-Aldrich or Fisher Scientific.

5.4.2. DNA manipulation

Plasmid cloning was done by Gibson Assembly Cloning Kit from New England Biolabs following the manufacturer's instructions or by the DNA assembler method[33]. The list of primers used in this chapter can be found in **Table 5.2**. For DNA manipulations, yeast plasmids were isolated using a Zymoprep II yeast plasmid isolation kit (Zymo Research, Irvine, CA) and transferred into *E. coli* for amplification. QIAprep Spin Plasmid Mini-prep Kits (Qiagen, Valencia, CA) were employed to prepare plasmid DNA from *E. coli*. All enzymes used for recombinant DNA cloning were from New England Biolabs unless otherwise noted. The products of PCR, digestion and ligation reactions were purified by QIAquick PCR Purification and Gel Extraction Kits (Qiagen, Valencia, CA).

The integration of the Cas9 gene was performed with the help of the pRS405 plasmid. Briefly, the P_{TEF1}-Cas9-T_{ADH2} or P_{ADH2}-Cas9-T_{ADH2} cassettes from a previous study[18] were subcloned into the multiple cloning site of the pRS405 plasmid. The recombinant pRS405 plasmids were then linearized in the middle of the *LEU2* marker, and transformed into the CAD strain.

The GFP-donor plasmid was constructed as follows. First, the *Bsm*BI site in the *URA3* maker of the pRS426 plasmid was removed with the primers PR361/PR362 using the Q5 Site-Directed Mutagenesis Kit from New England BioLabs, resulting in the construction of pRS426-mut. Then, a P_{SNR52}-*Bsm*BI-T_{SUP4} cassette for crRNA expression was PCR amplified with the primers PR364/PR365 from the pCRCT plasmid[18]. The crRNA cassette was then inserted between the two *Bsm*BI sites on the pRS426-mut backbone via Gibson assembly to create pRS426-mut-P_{SNR52}. Then the crRNA targeting the δ.a site was constructed through Golden Gate assembly method as previously described[18] with the primers PR346/PR347. The resultant

pRS426-mut-P_{SNR52}-crδ.a plasmid was then linearized by PCR with the primers PR436/PR437, and used for insertion of the homologous arms by Gibson assembly to create pRS426-mut-P_{SNR52}-crδ.a-LTRout. The homologous arms were synthesized as gBlocks Gene Fragment PR432 (Integrated DNA Technologies, Coralville, IO). Two δ.a sequences flanking the homologous arms and a *Pme*I site separating two arms were incorporated during DNA synthesis. The GFP expression cassette was amplified from the plasmid pRS405-TEF1p-GFP-PGK1t in **Chapter 3** with the primers PR451/PR452, and inserted at the *Pme*I site of pRS426-mut-P_{SNR52}-crδ.a-LTRout. The resultant plasmid pRS426-mut-P_{SNR52}-crδ.a-LTRout-P_{TEF1}-GFP-T_{PGK1} was linearized by PCR with the primers PR438/PR516, and used in Gibson assembly to be spliced with the *KanMX* cassette amplified from the pUG6 plasmid with the primers PR511/PR512, and the tracrRNA cassette amplified from the pRPR1p-tracrRNA plasmid[18] with the primers PR513/PR514. The final construct, pRS426-mut-P_{SNR52}-crδ.a-LTRout-P_{TEF1}-GFP-T_{PGK1}-KanMX-tracrRNA, was used as the GFP donor plasmid.

The same expression cassette in the RAGE2.0 library, P_{TEF1}-T_{PGK1}, was used in the integration donor, hence the construction of the donor plasmid library followed the same procedure described in **Section 4.4.3**. Briefly, the GFP donor plasmid was processed into backbone vectors for RAGE2.0 library construction, by replacing the *gfp* gene into the *ccdB* gene flanked by two 15 bp adaptor sequences with the primers PR608-PR611, as specified in **Section 4.4.3**.

5.4.3. RAGE3.0 cycle: transformation, induction and selection

The donor plasmid library was used to transform the CAD-P_{ADH2}-Cas9 strains by the standard LiAc/ssDNA/PEG protocol[34] with an optimized condition, where 50 μg plasmid DNA

was used to transform 40 OD_{600} unit competent yeast cells by heat shock at 42 °C for 1 hr. More than 10^6 independent yeast clones were obtained in each round using the optimized protocol. The yeast transformants were amplified on twenty 15 mm SC-U plates for 4 days. Cells were harvested from the SC-U plates, and 20 OD_{600} unit cells were spread on twenty 15 mm diameter YPAG plates containing 1 g/L G418. Cells were collected after 2 days, and 20 OD_{600} unit cells were spread on twenty 15 mm diameter SC-L plates containing 1 g/L 5-FOA (pH=4) and allowed for growth for 3 or 4 days. Then 20 OD_{600} unit harvested cells were inoculated into 20 mL SC-L liquid medium containing 1 g/L 5-FOA (pH=4) in a 125 mL baffled flask. After cultivation for 1 day, 10 OD_{600} unit cells were used to prepare competent cells in the SC-L medium for the next round of RAGE3.0.

The RAGE3.0 procedure can also be performed with liquid medium with some modifications. In particular, after transformation, all the cells were transferred into 50 mL SC-U medium to grow for 3 days. Then a re-inoculation step was included to minimize the presence of untransformed cells, by inoculating 20 mL fresh SC-U medium with 20 OD_{600} unit cells. After cultivated for 1 day, 20 OD_{600} unit cells were transferred into 20 mL YPAG medium containing 1 g/L G418, and induced for 2 days. For plasmid curation, 20 OD_{600} unit induced cells were resuspended in 20 mL SC-L medium with 1 g/L 5-FOA (pH=4) to grow until saturation. The cultivation in the SC+5-FOA medium was repeated once, and the cells were ready for the next round of RAGE3.0.

5.4.4. Analysis of the GFP reporter system

To ensure high expression level of Cas9, a procedure modified from the protocol in **Section 5.4.3** was used. Specifically, competent cells were prepared using the rich medium YPAD, and 5

μg GFP-donor plasmid was transformed into the yeast strains. For the negative control, 5 μg empty donor plasmid was used. The transformants were selected in the YPAD medium containing 1 g/L G418, and then automatically induced in the same selection medium when glucose was depleted. To obtain a high transformation efficiency, the same protocol in **Section 5.4.3** was used, where synthetic dropout medium was used for competent cell preparation and transformant selection, and induction was performed as a separate step. Different carbon sources, including 2% ethanol, 2% glucose and 2% galactose, were used in the induction YPA medium. In all cases, after plasmid curation, GFP fluorescence was analyzed by a LSR II Flow Cytometer (BD, Franklin Lakes, NJ).

5.4.5. Screen of glycerol-utilizing mutants

After three rounds of RAGE3.0, 10^7 cells were screened on twenty 15 mm diameter SC-L plates containing 3% glycerol as the sole carbon source. Twenty colonies with remarkably bigger sizes were selected for subsequent analysis. The yeast mutants were compared with the control strain (CAD-P_{ADH2}-Cas9) for glycerol utilization via an aerobic growth assay. Overnight culture in the SC-L 2% glucose medium was used to inoculate 3 ml SC-L 3% glycerol medium in 14 ml falcon tubes. The initial OD600 was set to 0.1, and the cell densities were monitored at 12 hr interval. The genomic DNAs of the G5, G6 and G10 strains were isolated by Wizard Genomic DNA Purification Kit (Promega, Madison, WI). Specific primer pairs, PR615/PR616, and PR617/PR618, were designed to PCR-amplify the integrated overexpression or knockdown cassettes, respectively. The PCR products were cloned into the empty donor plasmid via Gibson assembly, and individual *E. coli* colonies transformed with the Gibson assembly reaction mix

was picked to prepare the plasmids. DNA sequencing analysis was performed with the primers "S TEF1p For" and "S PGK1t Rev" in **Chapter 3**.

Table 5.2. List of primers used in this chapter

	Primer name	Primer sequence (5'→3')
PR361	p416-ura3-Mut For	ATGACAAGGGTGACGCATTGG
PR362	p416-ura3-Mut Rev	CTAAACCCACACCGGGTG
PR364	pRS-BsmBI-SNR52 For	ATCACGAGGCCCTTTTCTTTGAAAAGATAATGTATG
PR365	SUP4-BsmBI-pRS Rev	CAGACAAGCTGTGACAGACATAAAAAACAAAAAAAG
PR346	crDelta-2 F	AAACGAAGTTCTCCTCGAGGATATG
PR347	crDelta-2 R	AAAACATATCCTCGAGGAGAACTTC
PR432 (gBlock)	pRS-H3LTR-end-out (pRS-homolog-**δ.a**-arm δ1-*PmeI*-arm δ2-**δ.a**-pRS-homolog)	AATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG**GAA GTTCTCCTCGAGGATATAGG**TGTTGGAATAGAAATCAACTATC ATCTACTAACTAGTATTTACATTACTAGTATATTATCATATACGGT GTTAGAAGATGACGCAAATGATGAGAAATAGTCATCTAAATTAG TGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATCAATGA ATATAAACATATAAAATGATGATAATAATATTTATAGAATTGTGT AGAATTGCAGATTCCCTTTTATGGATTCCTAAATCCTTGAGGA*GT TTAAAC*GAACTTCTAGTATATTCTGTATACCTAATATTATAGCCTT TATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCATTT CTCA**GAAGTTCTCCTCGAGGATATAGG**TCACTGCCCGCTTTCCA GTCGGGAAACCTGTCGTGCCAGC
PR436	pRS-H3LTR-up Rev	CCTGATGCGGTATTTTCTCC
PR437	pRS-H3LTR-down For	CCAGTCGGGAAACCTGTCG
PR438	pRS-H3LTR-down Rev	CGACAGGTTTCCCGACTGG
PR451	LTR-out-TEF1p For	GATTCCTAAATCCTTGAGGAACACACCATAGCTTCAAAATG
PR452	LTR-out-PGK1t Rev	ACAGAATATACTAGAAGTTCCAGGAAGAATACACTATAC
PR511	crD2-KanMX For	CGGGAAACCTGTCGTGCCAGTAGGTCTAGAGATCTGTTTAGC
PR512	KanMX-Rpr1 Rev	GTTATGTTCAATTGGCAGATCATTAAGGGTTCTCGAGAGC
PR513	KanMX-Rpr1 For	GCTCTCGAGAACCCTTAATGATCTGCCAATTGAACATAAC
PR514	Rpr1t-pRS Rev	ATTCATTAATGCAGCCGCGGTCTTTCTGTATCGCAAATAAG
PR515	crD2-dKanMX For	CGGGAAACCTGTCGTGCCAGACATCCGAACATAAACAACC
PR516	Rpr1t-pRS For	CCGCGGCTGCATTAATG
PR615	TEF1p-15F For	AATCTAAGTTTTAATTACAAAAAGCAGTGG
PR616	15F-PGK1t Rev	TCGATTTCAATTCAATTCAATCGGGGTACG
PR617	TEF1p-15R For	AATCTAAGTTTTAATTACAAACGGGGTACG
PR618	15R-PGK1t Rev	TCGATTTCAATTCAATTCAATAAGCAGTGG

5.5. References

1. Linshiz, G. et al. PaR-PaR laboratory automation platform. *ACS Synth Biol* **2**, 216-222 (2013).
2. Esvelt, K.M., Carlson, J.C. & Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011).
3. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).
4. Si, T., Luo, Y., Bao, Z. & Zhao, H. RNAi-assisted genome evolution in *Saccharomyces cerevisiae* for complex phenotype engineering. *ACS Synth Biol*, in press (2014).
5. Storici, F., Durham, C.L., Gordenin, D.A. & Resnick, M.A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci U S A* **100**, 14994-14999 (2003).
6. DiCarlo, J.E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* **41**, 4336-4343 (2013).
7. Sakai, A., Shimizu, Y. & Hishinuma, F. Integration of heterologous genes into the chromosome of *Saccharomyces cerevisiae* using a delta sequence of yeast retrotransposon Ty. *Appl Microbiol Biotechnol* **33**, 302-306 (1990).
8. Parekh, R.N., Shaw, M.R. & Wittrup, K.D. An integrating vector for tunable, high copy, stable integration into the dispersed Ty delta sites of *Saccharomyces cerevisiae*. *Biotechnol Prog* **12**, 16-21 (1996).
9. Lee, F.W. & Da Silva, N.A. Improved efficiency and stability of multiple cloned gene insertions at the delta sequences of *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **48**, 339-345 (1997).

10. Hauber, J., Nelbock-Hochstetter, P. & Feldmann, H. Nucleotide sequence and characteristics of a Ty element from yeast. *Nucleic Acids Res* **13**, 2745-2758 (1985).

11. Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A. & Voytas, D.F. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* **8**, 464-478 (1998).

12. Lee, F.W. & Da Silva, N.A. Sequential delta-integration for the regulated insertion of cloned genes in *Saccharomyces cerevisiae*. *Biotechnol Prog* **13**, 368-373 (1997).

13. Chung, Y.-S. et al. Stable expression of xylose reductase gene enhances xylitol production in recombinant *Saccharomyces cerevisiae*. *Enzyme Microb Technol* **30**, 809-816 (2002).

14. Kim, Y.S., Kim, S.Y., Kim, J.H. & Kim, S.C. Xylitol production using recombinant *Saccharomyces cerevisiae* containing multiple xylose reductase genes at chromosomal delta-sequences. *J Biotechnol* **67**, 159-171 (1999).

15. Yamada, R. et al. Cocktail delta-integration: a novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. *Microb Cell Fact* **9**, 32 (2010).

16. Yuan, J. & Ching, C.B. Combinatorial assembly of large biochemical pathways into yeast chromosomes for improved production of value-added compounds. *ACS Synth Biol* (2014).

17. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

18. Zehua Bao, H.X., Jing Liang, Xiong Xiong, Ning Sun, Tong Si and & Zhao, H. A Homology Integrated CRISPR-Cas (HI-CRISPR) system for one-step multi-gene disruptions in *Saccharomyces cerevisiae*. *ACS Synth Biol*, submitted (2014).

19. Lee, K.M. & DaSilva, N.A. Evaluation of the *Saccharomyces cerevisiae ADH2* promoter for protein synthesis. *Yeast* **22**, 431-440 (2005).

20. Lynch, M.D., Warnecke, T. & Gill, R.T. SCALEs: multiscale analysis of library enrichment. *Nat Methods* **4**, 87-93 (2007).

21. Ho, C.H. et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol* **27**, 369-377 (2009).

22. Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391 (2002).

23. Winzeler, E.A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).

24. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).

25. Breslow, D.K. et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* **5**, 711-718 (2008).

26. Du, J., Yuan, Y., Si, T., Lian, J. & Zhao, H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res* **40**, e142 (2012).

27. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**, 1027-1032 (2006).

28. Tong, A.H. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364-2368 (2001).

29. Pan, X. et al. A robust toolkit for functional profiling of the yeast genome. *Mol Cell* **16**, 487-496 (2004).

30. Costanzo, M. et al. The genetic landscape of a cell. *Science* **327**, 425-431 (2010).

31. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* **9**, 258-267 (2007).

32. Alper, H., Moxley, J., Nevoigt, E., Fink, G.R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565-1568 (2006).

33. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**, e16 (2009).

34. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protocols* **2**, 31-34 (2007).