Using the Cre-loxP system to randomize target gene expression states and generate diverse phenotypes

by

Bradley Niesner

B.S. Chemical Engineering
Rice University, 2003

M.S. Chemical Engineering Practice
Massachusetts Institute of Technology, 2010

Submitted to the Department of Chemical Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering
at the
Massachusetts Institute of Technology

June 2013

Signature of Author: _____
Department of Chemical Engineering
May 13, 2013

Certified by: _____
Narendra Maheshri
Assistant Professor of Chemical Engineering
Thesis Supervisor

Accepted by: _____
Patrick S. Doyle
Professor of Chemical Engineering
Chairman, Committee of Graduate Students

Using the Cre-loxP system to randomize target gene expression states and generate diverse phenotypes

by

Bradley Niesner

Submitted to the Department of Chemical Engineering in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Chemical Engineering

Abstract

Modifying the expression of multiple genes enables both deeper understanding of their function and the engineering of complex multigenic cellular phenotypes. However, deletion or overexpression of multiple genes is typically laborious and involves multiple sequential genetic modifications. Here we describe a strategy to randomize the expression state of multiple genes in *S. cerevisiae* using Cre-loxP recombination. By inserting promoters flanked by inverted loxP sites in front of a gene of interest we can randomly alter its expression by turning it OFF or ON, or between 4 distinct expression states. We show at least 6 genes can be randomized independently and argue that using orthogonal loxP sites and an additional recombinase should increase this number to at least 30. Finally, we show how combining this strategy with mating allows easy introduction of native regulation as an additional expression state and use this to probe the role of 4 different enzymes involved in base excision repair in tolerance to methyl methane sulfonate (MMS), a genotoxic DNA alkylating agent. The set of vectors developed here can be used to randomize the expression of both heterologous genes and endogenous genes, and could immediately prove useful for metabolic engineering in yeast. Because Cre-loxP recombination works in many organisms, this strategy should be readily extendable.

Thesis Supervisor: Narendra Maheshri
Title: Assistant Professor of Chemical Engineering

Dedication

Dedicated to my family, who put up with my dinner table conversations about science because, deep down, they know that their love and support are the main reasons I always have so much to talk about.

Contents

Chapter 1: Techniques for changing the expression of target genes

(Portions of Chapters 1, 2, and 3 are adapted from Using the Cre-lox system to randomize target gene expression states and generate diverse phenotypes *by Bradley Niesner and Narendra Maheshri, accepted to Biotechnology and Bioengineering April 19, 2013)*

1.1 Complex genetic systems necessitate modifying the expression of multiple genes

Changing the transcriptional expression of genes of interest is a fundamental tool for understanding or engineering phenotypes in any organism. In the simplest cases, a single gene can be essential or have phenotypic consequences in a particular situation. However, many phenotypes depend on multiple genes whose products interact within a large network. This can lead to *epistasis*, where (a set of) gene(s) expression states can affect the expression of another (set of) gene(s), and have synergistic effects on a quantitative phenotype. Elucidating epistatic gene-gene interactions contributes to understanding the molecular mechanisms of multigenic phenotypes. For example, for the genetic interactions in Figure 1 (top), deletion of *GeneA* would result in decreased expression of *GeneC*, implying a positive genetic interaction between *Gene A* and *GeneC*. More complex epistatic relationships arise when two or more genes contribute to a quantitative phenotype in a way that is not simply additive. Consider the network in Figure 1 (bottom) that is responsible for a quantitative phenotype, such as production of a valuable metabolite. Both *GeneA* and *GeneB* are responsible for heightened production, but coexpression of *GeneA* and *GeneB* increases production beyond the individual effects of either one individually.

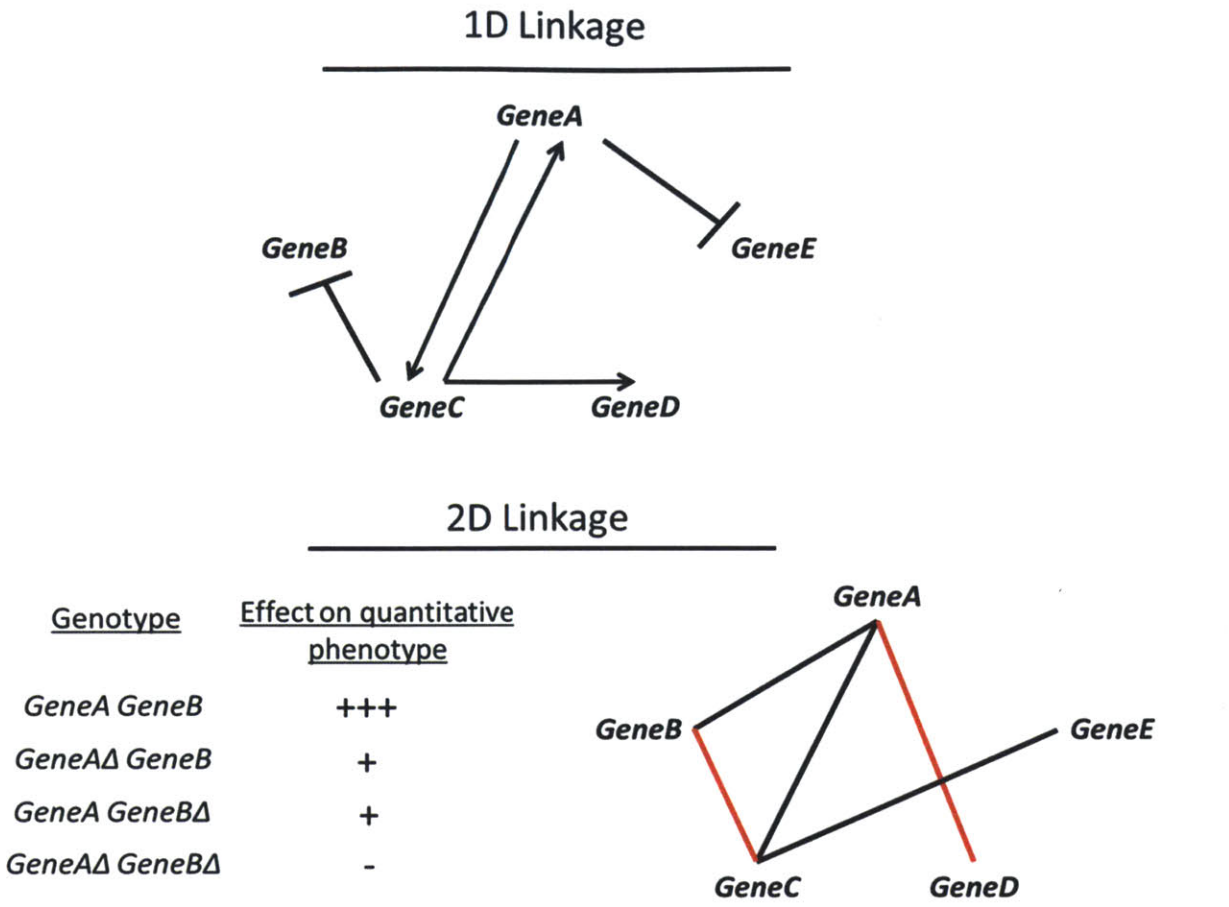


Figure 1: Examples of epistatic linkage. 1D epistatic linkage (top) implies that a gene affects the expression of another gene by increasing its expression when present (←) or decreasing it (⊦).2D epistatic linkage (bottom) implies that changing the expression of two genes creates an outcome that is not merely additive of the characteristics of changing each gene individually. Synergistic relationships are in black, antagonistic in red.

Global pairwise combinations of deletion mutants in yeast have yielded valuable insight into genome-wide epistatic relationships. Epistatic maps similar to Figure 1 (bottom, right) identify pairs within groups of related genes which have synergistic or antagonistic interactions and have been created for genes involved in chromosomal structure (Tong et al. 2004; Jasnos and Korona 2007), metabolism (Segre et al. 2004), and growth fitness (Jasnos and Korona 2007). Large-scale epistatic studies also demonstrate the underlying complexity of genetic systems: when 132 strains with single deletions of genes known to affect chromosomal stability, replication, and

division were crossed with the single-deletion library of all nonessential yeast genes (~4,000 genes) the results showed an average of 34 synergistic or antagonistic interactions affecting growth fitness per gene (Tong et al. 2004). Further, combining gene deletions known to cause growth defects in certain environments revealed that double deletions of genes in that same group generally *increased* fitness in the same growth conditions relative to single deletions, most likely due to synergistic or suppressive epistatic interactions (Jasnos and Korona 2007). These results indicate that yeast genetic networks have high degrees of interaction and redundancy, lending further urgency for the need to evaluate groups of genes in a combinatorial manner, even when only using deletion mutants.

As the number of genes under investigation increases, traditional techniques may prevent fully sampling the phenotypic landscape: the range of quantitative traits resulting from combinatorial expression engineering of underlying genes. The need to be thorough must be weighed against constraints of time, resources, and available assays. Traditional methods for genetic engineering rely on sequential genomic changes, with a single genetic target per step. Other than introducing genetic material on plasmids, integration of new sequences into a specific genomic locus relies on homologous recombination, a process which has low efficiency in even the most genetically tractable organisms. Each genetic change (deletion, overexpression, etc.) requires one step, but the number of *combinations* of each desired change means that the number of strains which must be constructed scales exponentially. Making all possible combinations of six gene deletions requires $2^6 = 64$ different strains. Each step also requires a selectable marker which can either be used only once or requires additional steps to remove and reuse. Marker removal can be accomplished by direct deletion if it is counter-selectable or via recombination if directed repeats flank the marker. Removing markers through recombination leaves a single

repeat scar in the genome, making further use of the same marker cassette or plasmid difficult due to sequence homology at many loci. These limitations allow the generation of expression libraries or engineered strains which do not require more than several sequential changes, but hamper in-depth epistatic studies among a group of target genes which may number more than available markers or require more strain construction time than is feasible. In the context of strain engineering, where the desired goal may be to generate *de novo* phenotypes or tune existing ones, it is desirable to not only eliminate target gene expression through deletion, but vary expression levels, further increasing the number of strains needed for a combinatorial expression library.

Changing the expression of more than two genes in individual strains has allowed in-depth analysis of phenotypic landscapes. Combining deletion and constitutive overexpression of target genes probed sources of robustness in the induction of the galactose utilization pathway in yeast and revealed a possible minimum feedback architecture for dosage invariance in signaling networks (Acar et al. 2010). Triple and even quadruple deletions among genes associated with DNA repair in yeast revealed genetic interactions whereby additional deletions suppress the detrimental effects of single deletion mutants (Boiteux and Guillet 2004). Metabolic networks are tightly regulated at the level of both transcription and enzyme activity, representing challenging targets for expression engineering. Various combinations of deletions known to respectively increase lycopene production in *E. coli* resulted in both increased and decreased production when crossed with each other (Alper, Miyaoku, and Stephanopoulos 2005) and a triple deletion of important regulators allowed a 41% increase in carbon flux through the yeast galactose utilization pathway (Ostergaard et al. 2000).

The need for genetic tools capable of generating expression diversity across multiple targets with minimal steps and/or markers has led to a range of innovative techniques (discussed in section 1.2) to accomplish this goal. In Chapter 2 we describe a method using Cre-catalyzed recombination to randomize promoter orientation and, therefore, expression levels at target loci. This method can affect endogenous as well as heterologous gene expression and rapidly generates individual yeast with different combinations of expression levels among the target genes. The method may extend beyond yeast because the Cre-catalyzed recombination is functional in organisms from bacteria to mammalian cells.

1.2 Methods for creating diverse expression libraries

Overcoming the limitations of traditional deletion or overexpression of many gene targets in combination generally requires increased efficiency of transformation (or transfection) and/or generation of multiple expression profiles in a population (expression diversity) with only one transformation step. Creating a library of combinatorial expression states across a population requires building exponentially increasing numbers of strains with each new genetic target and can easily become the limiting step in genetic studies. This section will summarize a number of methods designed to improve upon systematic, sequential introduction, deletion or modification of single genes and compare them in the context of applicability in different organisms, ability to affect heterologous and endogenous genes, and the ease of modifying multiple loci.

Diverse expression states through transformation of sequence libraries

Promoter libraries and error-prone PCR generate multiple genotypes at one target locus per transformation step. Promoter libraries can sample gene expression at many non-zero, discrete levels and have been used to engineer desirable phenotypes in yeast and bacteria (Alper et al. 2005; Jensen and Hammer 1998; Pfleger et al. 2006). Besides being the source of promoter

libraries, error-prone PCR is also commonly used to introduce mutations into open reading frames (ORFs) for directed evolution (Cirino, Mayer, and Umeno 2003) and can also include the promoter sequence in the amplified product (Cadwell and Joyce 1992). Traditional implementation of either method is generally limited to a single locus, sometimes with multiple rounds of selection and transformation in the case of error-prone PCR.

Titratable promoters allow rapid sampling of gene expression levels using small molecules and are advantageous because only one transformation step is necessary to insert the promoter. Combinatorial expression among target genes can be accomplished through different titratable promoters (Dueber et al. 2009), however such promoters are limited in number and small molecules, as well as inducible signals, may have pleiotropic effect (Cox, Surette, and Elowitz 2007; Guo et al. 2008; Maya et al. 2008).

Multiplex automated genome engineering (MAGE) uses a pool of synthetic single-stranded DNA (ssDNA) oligomers to generate genomic diversity in engineered *E. coli*. The ssDNA library contains oligomers with sequence homology to one or more target loci, along with random mismatches, insertions, and deletions. The λ-phage protein β facilitates binding between the oligomer and genomic DNA, and is expressed during sequential rounds of growth in the presence of the ssDNA library. MAGE was reported to have an allelic replacement efficiency of 30% over the course of 2.5 hours of growth in conditions designed to introduce the oligomers into the cells (Wang et al. 2009). Although the oligomers have sequence homology to their intended target locus, they generally anneal to the lagging strand of the replication fork due to its accessibility during replication (Ellis et al. 2001). As such, MAGE can target both endogenous and heterologous genes, provided the heterologous genes are introduced.

High integration efficiency allows replacement at many small genomic targets using MAGE. Successive rounds of growth and introduction of DNA oligomers rapidly generates a heterogeneous population with the possibility that every cell has diverged from the ancestral genotype after several rounds. As a demonstration, the authors used MAGE to target the ribosomal binding sequences of 24 genes associated with lycopene biosynthesis in order to change the translation efficiency of transcribed mRNAs, effectively changing their expression levels. A number of isolated clones demonstrated increases in lycopene production over the ancestral strain and had changes at up to five of the target loci. The time scale required for changes at so many loci is dramatically lower for MAGE than traditional methods, most likely due to the high efficiency of modification.

Major hurdles exist for extending MAGE to other organisms. The ease of introducing heterogeneity at targeted locations in the genome using MAGE takes advantage of the high growth rate and access to the genome which is characteristic of exponentially growing bacteria. With replication occurring once every 20 to 30 minutes, lagging strand accessibility is 3- to 4-fold higher than simple eukaryotes such as yeast, with higher organisms replicating much more slowly, although yeast can be transformed by ssDNA oligomers (Moerschell, Tsunasawa, and Sherman 1988). The upper limit on the number of possible target loci is influenced by both the integration efficency and the total length of the target loci. The authors targeted short ribosomal binding sites (RBSs) in order to overcome the disadvantage that the oligomers lose efficiency quickly if their length increases or decreases from 90 bp. This technique could also be applied in higher organisms by targeting short promoter elements that affect transcription, such as TATA boxes, to cover as many genes as in bacteria. In terms of an upper limit on target loci, the oligomer library could be expanded to include more loci; however that would limit the coverage

of the technique by increasing the required size of the oligomer pool and decreasing the chance of individual cells acquiring changes at a large fraction of the targets. Even with an efficiency of 30%, it would require 26 rounds of integration for 10% of the population to have accumulated 10 successful integration events. It is also important to account for the number of different sequences targeted to each locus. With 4 different sequences per locus, 10 target loci, and a population of 10^7 cells the fraction of the population which has accumulated 10 integration events after 26 rounds of integration (10^6 cells) still only represents a 1X coverage of possible genotypes (4^{10}~10^6). While a population size of 10^7 could be increased 100-fold for unicellular microbes such as bacteria and yeast, it represents a rough practical limit for plant or mammalian cell culture. The limit on oligo size also means that the total genomic *sequence* which can be targeted is on the order of several kilobases (kb). With 24 targets and 100 bp oligos, the effective target sequence length is only 2.4 kb. This would decrease the ability of MAGE to mutate multiple promoters or ORFs as these are also on the order of kilobases in length in contrast to the short elements used by the authors.

Random ligation or homologous recombination of gene pools

Using a pool of gene ORFs, combinatorial libraries can be generated by randomly incorporating gene fragments through ligation or recombination, with eventual products possessing zero to multiple copies of each gene. Based on the length of the assembled DNA, it can be introduced through direct genomic integration or the use of an artificial chromosome. The efficiency of homologous recombination in *S. cerevisiae* allows *in vivo* construction of multiple DNA fragments (Raymond, Pownder, and Sexson 1999) and applies to large product sequences meant for eventual genomic integration (Shao, Zhao, and Zhao 2009). Another method of assembly deployable in a larger range of organisms is the *in vitro* ligation of pathway components

followed by its introduction on an artificial chromosome. Random ligation of components of the flavonoid biosynthetic pathway, followed by screening, resulted in the creation of novel strains that could produce a variety of different flavonoids in yeast (Naesby et al. 2009). Large artificial chromosomes have been used in bacteria (Shizuya et al. 1992) and mammalian cells (Henning et al. 1999), in addition to yeast, and require only one transformation/transfection step to introduce into an organism, but the method is poorly suited to studies of groups of endogenous genes.

Enhanced homologous recombination through break-induced repair

Difficulties of targeted genome modifications in metazoans stem from the inefficiency of homologous recombination. Because homologous recombination of donor DNA at a particular locus requires a double-strand break (DSB) at that locus, the major strategy to increase its efficiency is by the targeted introduction of a DSB. For example, in yeast introduction of a DSB using a homing endonuclease increases transformation efficiencies by over three orders of magnitude (Storici et al. 2003).

Expression of zinc finger endonucleases has aided in strand integration at target loci, but a new method of endonuclease-catalyzed breaks has increased the efficiency of genome modification. Expression of the bacterial Cas9 endonuclease in cells exposed to RNAs characterized by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and sequence homology to an endogenous locus catalyze DSBs at that locus. This site-specific DSB generation creates the expected increase in double- and single-strand DNA integration. Controlled expression of Cas9 along with guide RNAs (gRNAs) to aid in genomic modifications has been conducted in a wide range of organisms (Cong et al. 2013; DiCarlo et al. 2013).

The CRISPR method can also be multiplexed. Adding multiple targeting sequences to the gRNA molecule or introducing multiple gRNA molecules with different sequences can generate DSBs at multiple loci. A single gRNA molecule targeting two sequences can induce DSB formation at both loci in ~1% of a population (Cong et al. 2013). .

The predominant drawback to the CRIPSR method is that efficiency of arbitrary sequence integration may not be high enough for multiplexed strain engineering. Successful integration efficiencies of up to 10% have been demonstrated in human cell lines (Mali et al. 2013) and yeast (DiCarlo et al. 2013) although DSB generation is also dependent on both the gRNA target sequence and the type of cell line in the case of human cells. Even at 10% efficiency, 7 rounds of transformation/transfection would be required before 10% of the population experiences 2 successful integration events. In each round, the fraction of the population which has experienced multiple integration events drops off exponentially. CRISPR efficiencies may improve rapidly in model organisms like yeast with highly efficient HR, but higher organisms may pose a problem, possibly limiting CRIPSR to a highly efficient transformation method, but one which still requires sequential steps. Improvements in efficiency could allow targeting many genes with a DNA library, similar to MAGE, but with more potential to work in higher organisms.

Modifying expression levels of multiple genes with recombinases

Site-specific recombinases are enzymes that recombine two short (<50bp), oriented target sequences. They are used by phage to integrate genomic material into a host chromosome or have other roles in phage genome maintenance (Van Duyne 2001; Reynolds, Murray, and Szostak 1987) and a range of different recombinases have been purified from bacteria and yeast. Due to their small size, the recombination target sequences have little chance of recombining in

the absence of the recombinase. Historically, recombinases have been used to disrupt (Rohlmann et al. 1996) or activate (O'Gorman, Fox, and Wahl 1991) genes in host organisms, even somatic cells, based on clever arrangement of recombinase-specific sequences in a cassette. Recombinase-based disruption is especially useful for affecting gene expression at particular stages of development, as target genes may be essential for embryo development, but non-essential later in life.

Two widely used recombinases are Flp, from yeast, and Cre, from bacteriophage P1. They target the sequences known as FRT and loxP, respectively, and should function in nearly any organism (Sauer 1994; Nagy 2000). Recombinases with similar activity but different sequence specificity have been engineered (Hartung and Kisters-Woike 1998) or isolated from other organisms (Christ, Corona, and Droge 2002; Ringrose, Angrand, and Stewart 1997). The range of available recombinases allows temporal control of activity and specific targeting of loci although, as with titratable promoters, inducible systems to control recombinase expression are limited in number and sequential genomic modification must be conducted to insert a cassette to be targeted by the recombinase.

1.3 General limitations of current modification methods

Efforts to improve the diversity of expression levels for a set of target genes in a population generally involve homologous recombination of diverse genetic libraries. In the case of recombinase-based promoter inversion the generation of expression diversity occurs at every locus after sequential transformation steps, meaning that transformation efficiency does not influence the fraction of loci which are affected. Our recombination system can also be used with both endogenous and heterologous genes, and should extent to organisms other than yeast. Table 1 summarizes the attributes of the techniques discussed in Section 1.2.

A unique advantage of recombinase-based disruption is the ability to target genes which are essential in certain stages of development. Cassettes can be designed so that the gene of interest is active until the recombinase is expressed. Despite the multiplex capability of MAGE and the potential to conduct multiplex targeting with CRISPR, it is unclear if their integration events ever occur in the same replication phase, or in subsequent phases within a single round of exposure to targeted integration sequences. If integration occurs in different replication events, these techniques may not be able to delete or decrease expression of essential genes. Disrupting a *group* of genes in a single round of recombination, however, could reveal suppressor deletion combinations among essential genes which are actually not lethal.

Table 1: Summary of techniques used to generate expression diversity.

Technique	Ability to affect multiple gene targets	Deployable in different organisms	Endogenous or heterologous gene targets
Promoter library	No, but can sequentially target different loci	Yes	Both
Titratable promoter	Yes, demonstrated with 2 separate promoters, but limited in number and with possible pleiotropic effects	Yes	Both
Random ligation of ORFs	Yes, demonstrated with >10 genes	Yes	Easily deployable with heterologous genes, would require sequential deletion of endogenous genes
MAGE	Yes, sequence modifications documented at up to 5 loci out of a pool of 24 targets after multiple rounds of transformation	Only demonstrated in bacteria	Both
CRISPR-Cas9	In theory, but has not been demonstrated	Yes	Both
Recombinase-based promoter recombination	Yes, demonstrated with 6 target genes, and in principle **n** target genes randomized after **n** sequential transformations	Yes	Both

In Chapter 2 we describe a strategy to randomize the expression levels of both native and heterologous genes in *S. cerevisiae* with a limited number of required genetic modifications by using the Cre/loxP system. We develop and test vectors that include both a 2-state and 4-state promoter design where the downstream gene can adopt either 2 (OFF and ON) or 4 (OFF, LO, INT, HI) constitutive expression states after 'randomization' with Cre expression. By introducing one randomizable promoter per gene, a range of discrete expression states is generated in a previously isogenic population upon induction of Cre. Up to 5 copies of a 2-state promoter, and 2 copies of a 4-state promoter are reliably randomized without significant loss of the cassette. Additional genes can be added by using two sets of orthogonal loxP sites, which allowed 6 copies of a 2-state promoter and could in principle be extended to 30 copies, discussed in Chapter 4. In contrast to the methods discussed previously, our strategy generates expression diversity after a series of sequential transformations introduce randomizable loci into the population.

Furthermore, in Chapter 3 we demonstrate how combining a simple mating strategy and the 2-state promoters can randomize the expression state of up to 5 different genes involved in base excision repair between an OFF, ON, and natively regulated state. By screening the resulting library for growth in the presence of methyl methane sulfonate (MMS) – a genotoxic agent that alkylates DNA – we recover a set of genotypes which confer tolerance against this stress as well as demonstrate the ability of the promoter recombination system to generate phenotypic diversity.

1.4 References

Acar, Murat, Bernardo F. Pando, Frances H. Arnold, Michael B. Elowitz, and Alexander van Oudenaarden. 2010. "A General Mechanism for Network-Dosage Compensation in Gene Circuits." *Science* 329 (5999): 1656-1660.

Alper, Hal, Curt Fischer, Elke Nevoigt, and Gregory Stephanopoulos. 2005. "Tuning Genetic Control through Promoter Engineering." *Proceedings of the National Academy of Sciences of the United States of America* 102 (36): 12678-12683.

Alper, Hal, Kohei Miyaoku, and Gregory Stephanopoulos. 2005. "Construction of Lycopene-Overproducing E. Coli Strains by Combining Systematic and Combinatorial Gene Knockout Targets." *Nature Biotechnology* 23 (5): 612.

Boiteux, Serge and Marie Guillet. 2004. "Abasic Sites in DNA: Repair and Biological Consequences in Saccharomyces Cerevisiae." *DNA Repair* 3 (1): 1-12.

Cadwell, R. Craig and Gerald F. Joyce. 1992. "Randomization of Genes by PCR Mutagenesis." *Genome Research* 2 (1): 28-33.

Christ, Nicole, Teresa Corona, and Peter Droge. 2002. "Site-Specific Recombination in Eukaryotic Cells Mediated by Mutant L Integrases: Implications for Synaptic Complex Formation and the Reactivity of Episomal DNA Segments." *Journal of Molecular Biology* 319 (2): 305-314.

Cirino, Patrick C., Kimberly M. Mayer, and Daisuke Umeno. 2003. "Generating Mutant Libraries using Error-Prone PCR." In *Directed Evolution Library Creation*, 3-9: Springer.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex Genome Engineering using CRISPR/Cas Systems." *Science* 339 (6121): 819-823.

Cox, Robert Sidney, Michael G. Surette, and Michael B. Elowitz. 2007. "Programming Gene Expression with Combinatorial Promoters." *Mol Syst Biol* 3.

DiCarlo, James E., Julie E. Norville, Prashant Mali, Xavier Rios, John Aach, and George M. Church. 2013. "Genome Engineering in Saccharomyces Cerevisiae using CRISPR-Cas Systems." *Nucleic Acids Research.*

Dueber, John E., Gabriel C. Wu, G. Reza Malmirchegini, Tae Seok Moon, Christopher J. Petzold, Adeeti V. Ullal, Kristala LJ Prather, and Jay D. Keasling. 2009. "Synthetic Protein Scaffolds Provide Modular Control Over Metabolic Flux." *Nature Biotechnology* 27 (8): 753-759.

Ellis, Hilary M., Daiguan Yu, Tina DiTizio, and Donald L. Court. 2001. "High Efficiency Mutagenesis, Repair, and Engineering of Chromosomal DNA using Single-Stranded Oligonucleotides." *Proceedings of the National Academy of Sciences* 98 (12): 6742-6746.

Guo, Z. Sheng, Qiang Li, David L. Bartlett, James Y. Yang, and Bingliang Fang. 2008. "Gene Transfer: The Challenge of Regulated Gene Expression." *Trends in Molecular Medicine* 14 (9): 410-418.

Hartung, Markus and Brigitte Kisters-Woike. 1998. "Cre Mutants with Altered DNA Binding Properties." *Journal of Biological Chemistry* 273 (36): 22884-22891.

Henning, Karla A., Elizabeth A. Novotny, Sheila T. Compton, Xin-Yuan Guan, Pu P. Liu, and Melissa A. Ashlock. 1999. "Human Artificial Chromosomes Generated by Modification of a Yeast Artificial Chromosome Containing both Human Alpha Satellite and Single-Copy DNA Sequences." *Proceedings of the National Academy of Sciences* 96 (2): 592-597.

Jasnos, Lukasz and Ryszard Korona. 2007. "Epistatic Buffering of Fitness Loss in Yeast Double Deletion Strains." *Nature Genetics* 39 (4): 550-554.

Jensen, Peter Ruhdal and Karin Hammer. 1998. "The Sequence of Spacers between the Consensus Sequences Modulates the Strength of Prokaryotic Promoters." *Applied and Environmental Microbiology* 64 (1): 82-87.

Mali, Prashant, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, and George M. Church. 2013. "RNA-Guided Human Genome Engineering Via Cas9." *Science* 339 (6121): 823-826.

Maya, Douglas, María José Quintero, de la Cruz Muñoz-Centeno, María, and Sebastián Chávez. 2008. "Systems for Applied Gene Control in Saccharomyces Cerevisiae." *Biotechnology Letters* 30 (6): 979-987.

Moerschell, R. P., S. Tsunasawa, and F. Sherman. 1988. "Transformation of Yeast with Synthetic Oligonucleotides." *Proceedings of the National Academy of Sciences* 85 (2): 524-528.

Naesby, Michael, SV Nielsen, CA Nielsen, Trine Green, Thomas Ø Tange, Ernesto Simón, Philipp Knechtle, Anders Hansson, Markus S. Schwab, and Olca Titiz. 2009. "Yeast Artificial Chromosomes Employed for Random Assembly of Biosynthetic Pathways and Production of Diverse Compounds in Saccharomyces Cerevisiae." *Microbial Cell Factories* 8 (1): 45.

Nagy, Andras. 2000. "Cre Recombinase: The Universal Reagent for Genome Tailoring." *Genesis* 26 (2): 99-109.

O'Gorman, STEPHEN, DANIEL T. Fox, and GEOFFREY M. Wahl. 1991. "Recombinase-Mediated Gene Activation and Site-Specific Integration in Mammalian Cells." *Science* 251 (4999): 1351-1355.

Ostergaard, Simon, Lisbeth Olsson, Mark Johnston, and Jens Nielsen. 2000. "Increasing Galactose Consumption by Saccharomyces Cerevisiae through Metabolic Engineering of the GAL Gene Regulatory Network." *Nature Biotechnology* 18 (12): 1283-1286.

Pfleger, Brian F., Douglas J. Pitera, Christina D. Smolke, and Jay D. Keasling. 2006. "Combinatorial Engineering of Intergenic Regions in Operons Tunes Expression of Multiple Genes." *Nature Biotechnology* 24 (8): 1027-1032.

Raymond, Christopher K., Tracey A. Pownder, and Shannon L. Sexson. 1999. "General Method for Plasmid Construction using Homologous Recombination." *BioTechniques* 26: 134-141.

Reynolds, AE, AW Murray, and JW Szostak. 1987. "Roles of the 2 Microns Gene Products in Stable Maintenance of the 2 Microns Plasmid of Saccharomyces Cerevisiae." *Molecular and Cellular Biology* 7 (10): 3566-3573.

Ringrose, Leonie, Pierre-Olivier Angrand, and A. Francis Stewart. 1997. "The Kw Recombinase, an Integrase from Kluyveromyces Waltii." *European Journal of Biochemistry* 248 (3): 903-912.

Rohlmann, Astrid, Michael Gotthardt, Thomas E. Willnow, Robert E. Hammer, and Joachim Herz. 1996. "Sustained Somatic Gene Inactivation by Viral Transfer of Cre Recombinase." *Nature Biotechnology* 14 (11): 1562-1565.

Sauer, Brian. 1994. "Site-Specific Recombination: Developments and Applications." *Current Opinion in Biotechnology* 5 (5): 521-527.

Segre, Daniel, Alexander DeLuna, George M. Church, and Roy Kishony. 2004. "Modular Epistasis in Yeast Metabolism." *Nature Genetics* 37 (1): 77-83.

Shao, Zengyi, Hua Zhao, and Huimin Zhao. 2009. "DNA Assembler, an in Vivo Genetic Method for Rapid Construction of Biochemical Pathways." *Nucleic Acids Research* 37 (2): e16-e16.

Shizuya, Hiroaki, Bruce Birren, Ung-Jin Kim, Valeria Mancino, Tatiana Slepak, Yoshiaki Tachiiri, and Melvin Simon. 1992. "Cloning and Stable Maintenance of 300-Kilobase-Pair Fragments of Human DNA in Escherichia Coli using an F-Factor-Based Vector." *Proceedings of the National Academy of Sciences* 89 (18): 8794-8797.

Storici, Francesca, Christopher L. Durham, Dmitry A. Gordenin, and Michael A. Resnick. 2003. "Chromosomal Site-Specific Double-Strand Breaks are Efficiently Targeted for Repair by Oligonucleotides in Yeast." *Proceedings of the National Academy of Sciences* 100 (25): 14994-14999.

Tong, Amy Hin Yan, Guillaume Lesage, Gary D. Bader, Huiming Ding, Hong Xu, Xiaofeng Xin, James Young, Gabriel F. Berriz, Renee L. Brost, and Michael Chang. 2004. "Global Mapping of the Yeast Genetic Interaction Network." *Science Signalling* 303 (5659): 808.

Van Duyne, Gregory D. 2001. "A Structural View of Cre-Loxp Site-Specific Recombination." *Annual Review of Biophysics and Biomolecular Structure* 30 (1): 87-104.

Wang, Harris H., Farren J. Isaacs, Peter A. Carr, Zachary Z. Sun, George Xu, Craig R. Forest, and George M. Church. 2009. "Programming Cells by Multiplex Genome Engineering and Accelerated Evolution." *Nature* 460 (7257): 894-898.

Chapter 2: Controlled promoter recombination using the Cre-loxP system

2.1 Cre-lox recombination

As discussed in Section 1.4, the Cre recombinase is one of two widely used recombinases in the field of molecular biology. It can be expressed in an organism to recombine two loxP sites: 34 bp sequences with two palindromic motifs separated by a directional 8 bp spacer. Depending on the strand of the loxP site and the directionality of the spacer region, recombination can result in irreversible excision or reversible inversion of the sequence between the loxP sites (Nagy 2000) (Figure 2). Sequences flanked by loxP sites will be referred to as 'floxed' sequences.

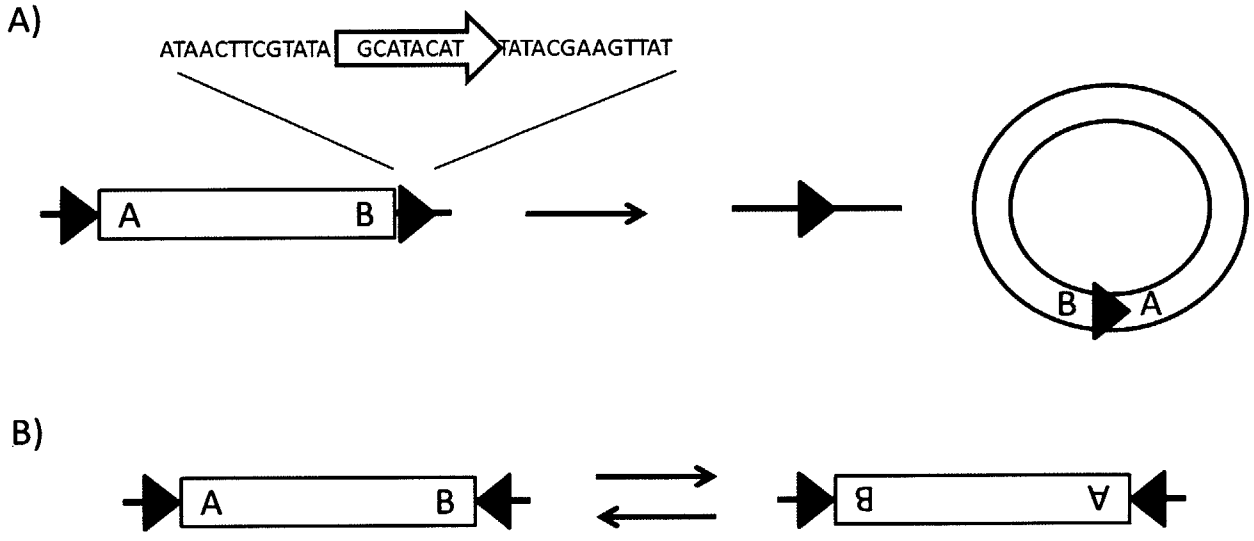


Figure 2: Cre-mediated recombination events on sequences flanked by loxP sites. LoxP sites are symbolized by black triangles and oriented according to the direction of the 8 bp spacer sequence within the site. A) The sequence of the loxP site with directionality of the central spacer, flanked by palindromic sequences. Recombination of parallel sites results in excision of the floxed sequence, with one loxP site remaining at the original locus. B) Recombination between loxP sites with opposite orientation results in reversible inversion of the floxed sequence.

Many variants of the loxP sequence have been engineered, most of which can be recombined by Cre. Most of these variant lox sequences can be recombined with other variants in the presence

of Cre but two variants, referred to as loxN and lox2272, are orthologous, meaning they can recombine with themselves, but cannot be recombined with each other or with loxP (Lee and Saito 1998).

2.2 A plasmid library for introducing floxed promoters in *S. cerevisiae*

We reasoned that the orientation of a constitutive promoter flanked by inverted loxP sites would effectively be randomized after a period of continuous Cre expression since recombination is fast and reversible. In one orientation, the promoter would drive the expression of the downstream gene and, when inverted, would effectively cease expression of the gene. Figure 3 illustrates this ON/OFF '2-state' promoter.

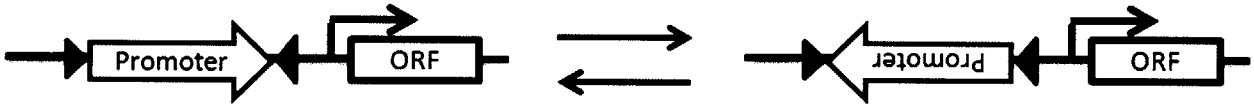


Figure 3: Two-state promoter recombination. Recombination by Cre reversibly inverts the promoter between an 'ON' state (left) and an 'OFF' state (right).

We constructed a series of yeast integration plasmids for building strains with target genes controlled by floxed promoters (Figure 4). The first type of vector, based on the pRS series (Sikorski and Hieter 1989), contains the promoter region of the constitutive yeast genes *TEF1*, *ACT1*, *ADH1*, or *MYO2*, driving expression of a fluorescent reporter. A heterologous gene can be cloned into the *EcoRI/NotI* restriction backbone in place of the reporter. The entire cassette can then be integrated into a specific marker locus by digestion inside the plasmid marker ORF. The second type of vector is based on the pFα6 disruption series (Longtine et al. 1998) and can be used to place a floxed promoter upstream of any endogenous gene. Primers can be designed to amplify a cassette containing a selectable marker followed by one of two different floxed

promoters (P_{TEF1} and P_{ACT1}) for integration and replacement of the native promoter. For each plasmid, the promoter is originally oriented so that the downstream gene is 'ON.'

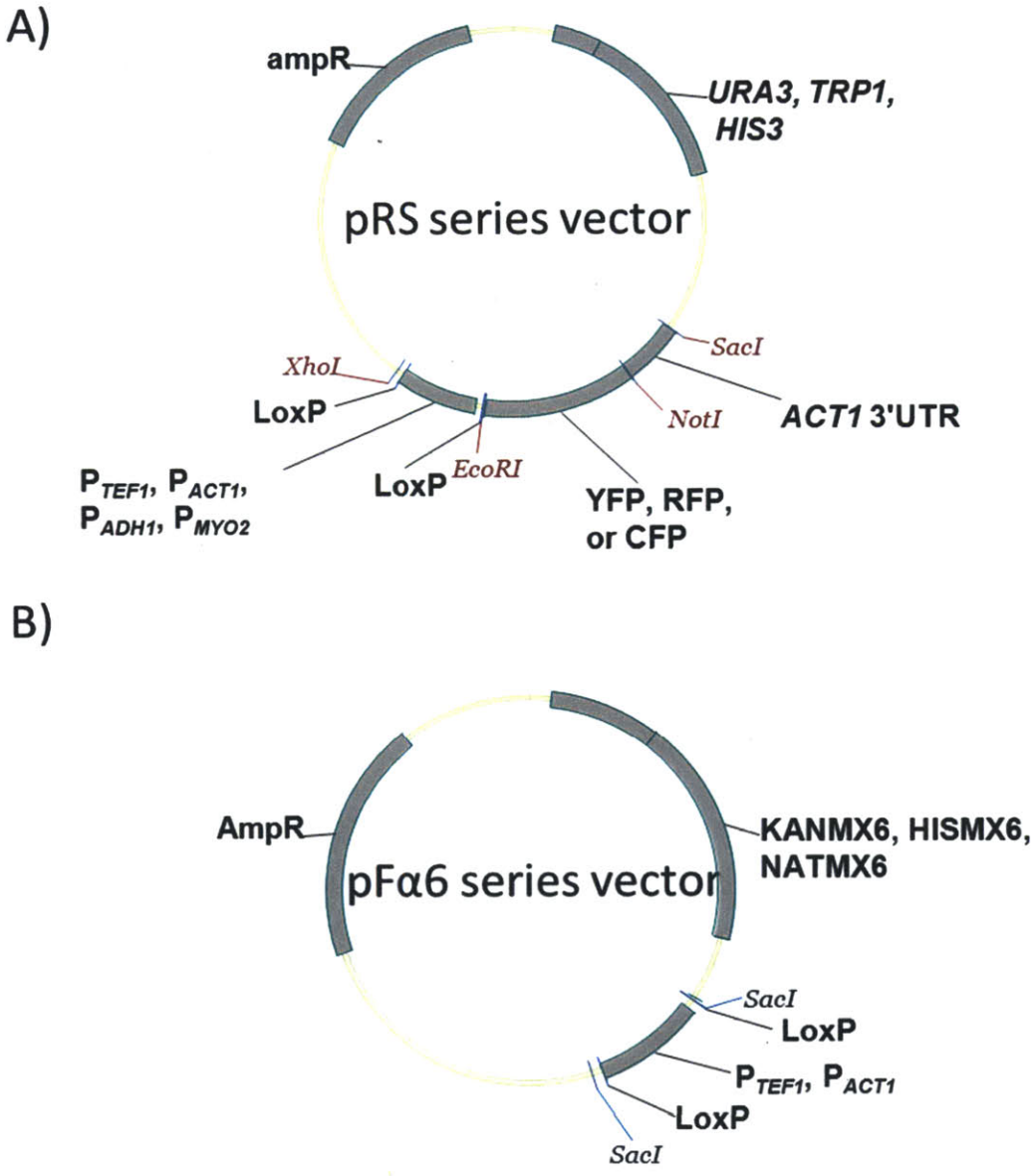


Figure 4: Vectors for introduction of floxed promoters in yeast. Promoters, downstream genes, markers, and loxP sites were varied to generate a set of vectors adaptable to randomize any gene's expression. A) A set of vectors to be integrated at the respective auxotrophic marker locus after restriction within the marker ORF. The floxed promoters and reporter can be excised and swapped using *EcoRI/XhoI* and *EcoRI/NotI* sites, respectively. B)

A set of PCR-amplifiable disruption vectors with non-homologous autotrophic or resistance makers. The marker and floxed promoter can be amplified using the sequences 5' – TGACGTGCGCAGCTCAGGGG – 3', and 5' – ATCGATGAATTCGAGCTC – 3' on the 3' end of sense and antisense primers, respectively.

2.3 Controlled expression of Cre recombinase

Promoter recombination requires a period of Cre expression in order to randomize orientation. However, tight control over Cre expression is desirable in order to prevent promiscuous recombination events. We obtained a centromeric plasmid with a galactose-inducible *GAL1* promoter driving Cre expression (Sauer 1987). Yeast activate the galactose utilizing pathway in the presence of galactose and absence of glucose, thus glucose inhibits galactose pathway activation (Adams 1972).

The protocol for expression of Cre is summarized in Table 2. A population of cells containing the galactose-inducible Cre plasmid and one or more integrated floxed promoters was grown in 2% raffinose media overnight in order to de-repress the galactose pathway. The initial orientation of the floxed promoters in all cells was ON. The population was then inoculated in 2% galactose media at an optical density (OD_{600}) of ~0.1. During this phase all cells activate the galactose pathway and express Cre. Each floxed promoter is inverted many times. Finally, the cells were inoculated in 2% glucose at an OD_{600} ~0.05. This low inoculation density allowed the cells to double more than five times, diluting out any Cre which remained and allowed the gene controlled by the floxed promoter to reach a stready-state expression level determined by the final orientation of the promoter.

Table 2: Cre expression protocol

Media[a]	Duration	Cre expression	Promoter orientation
2% raffinose	12-18 hr (overnight)	De-repressed	ON
2% galactose	6-12 hr	Expressed	Actively inverting
2% glucose	18-24 hr	Repressed	Fixed, ON or OFF

[a]All media are synthetic with leucine dropout to maintain the *LEU2*-marked Cre plasmid

2.4 Expression of Cre randomizes 2-state promoter orientation

To test the randomization strategy outlined in Section 2.3, we integrated a floxed *TEF1* promoter driving a yellow fluorescent protein reporter (P_{TEF1}-YFP) at the *HIS3* locus. We evaluated population-wide reporter expression via flow cytometry in strains containing just P_{TEF1}-YFP, P_{TEF1}-YFP and the Cre plasmid both before and after growth in galactose. A base strain containing neither the reporter nor the plasmid was used as the autofluorescence control. In Figure 5, analysis of YFP expression in the population containing both the floxed promoter and the Cre plasmid indicates that large fractions of the population are found in both the ON (35%) or OFF (65%) state for the promoter after growth in galactose media. Expression in the OFF state is indistinguishable from the control strain lacking the fluorescence reporter. We used a PCR strategy to verify promoter inversion had occurred (Figure 6). Importantly, randomization was dependent on growth in galactose, alleviating any concerns about leaky Cre expression when grown in 2% raffinose, when the galactose pathway is de-repressed.

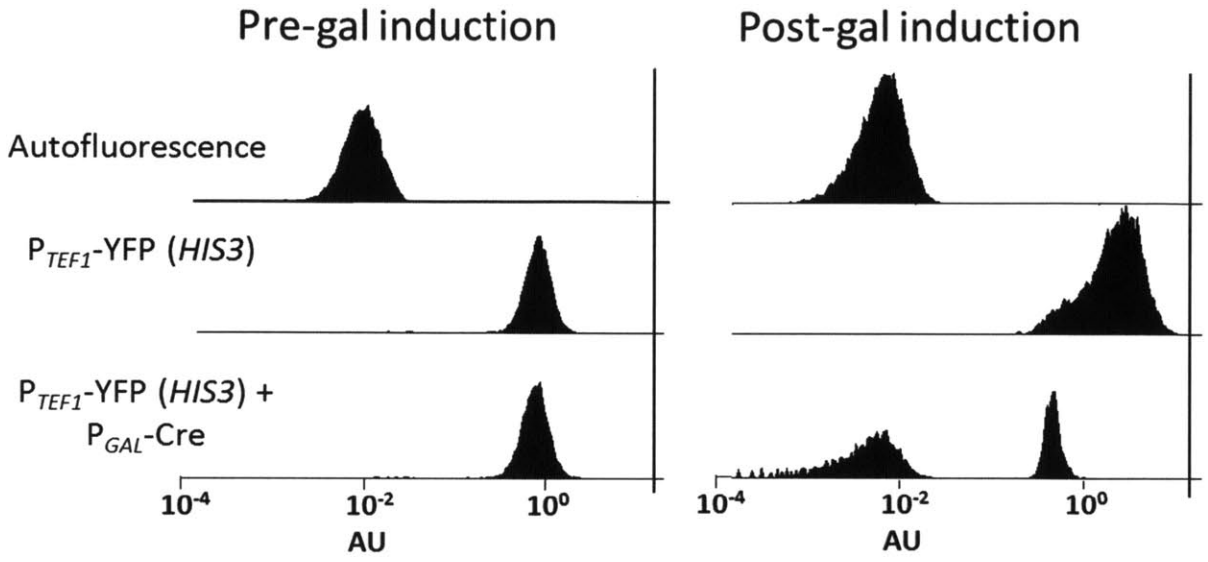


Figure 5: Cre-dependent promoter randomization. Fluorescence histograms of a floxed *TEF1* promoter driving YFP, *HIS3*::P_{TEF1}-YFP. All cells (*left*) begin in the ON state before Cre expression (*right*) but randomly settle in an ON or OFF state after Cre induction. Expression is normalized to the mean of the pre-induction control lacking the recombinase plasmid. N = 10,000 cells per plot.

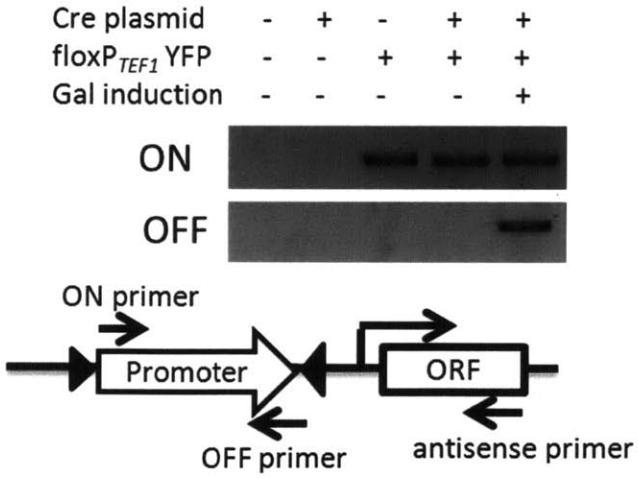


Figure 6: Molecular inversion of the floxed two-state promoter. Genomic PCR of multiple strains using the given primer configuration (bottom) shows the promoters randomize only after Cre induction.

While Cre is highly efficient, multiple inversion events may result in irreversible loss of the floxed promoter or recombination with other genomic regions leading to a large scale rearrangement. Moreover, unequal recombination between sister chromatids with inverted loxP sites can lead to chromosomal loss and cell death in mice (Lewandoski and Martin 1997; Grégoire and Kmita 2008). In fact, individual clones chosen after randomization whose promoters were in the ON and OFF states could be randomized again; the resulting phenotypic distribution was identical regardless of the initial state (Figure 7). In each case, the final ON population was similar to the 35% see for the first round of randomization. Still, we sought to minimize the time of Cre expression so as to reduce the chance of irreversible rearrangements. Therefore, we repeated the randomization but varied the amount of time the cells spent growing in galactose. Increasing the period of Cre expression to 6 hours increased the eventual fraction of cells in the OFF state, beyond which there was no change (Figure 8). Before 3 hours, it appears that a portion of the population has either not yet activated the galactose-inducible promoter or has not expressed enough Cre to generate recombination events.

Interestingly, we never saw a 50%/50% split in ON and OFF cells after randomization; instead, we observed a bias towards OFF cells. This bias was not due to growth differences in cells with the different expression states and depended on the genomic location where the floxed *TEF1* promoter was integrated: the OFF state fraction after 6 hours of galactose induction was 70(±3)%, 75(±2)%, and 88(±2)% at the *URA3*, *HIS3*, and *TRP1* loci respectively (n=3 replicates). While we do not know the origin of this bias, perhaps the forward and backward reaction rates of the inversion depend on either adjacent sequence or the regional genomic environment. Nevertheless, >10% of the population was always found to be ON post-randomization. Another source of the bias could be irreversible recombination events which

somehow excise the floxed promoter or render it inactive. The bias towards the OFF state supports this idea, as cells which are OFF are indistinguishable from cells which do not contain the reporter. The OFF population may contain a subset of cells which have experienced an irreversible recombination event. Section 2.5 addresses loss of floxed promoters in more detail.

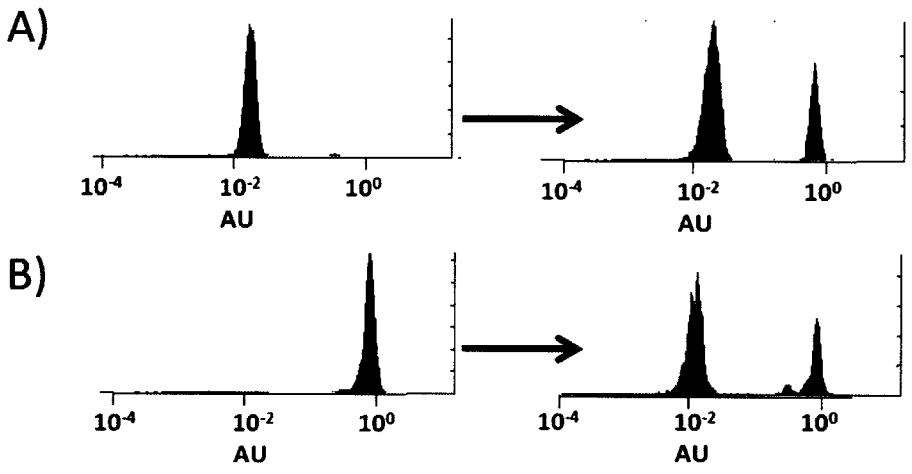


Figure 7: Reversibility of clones taken from a randomized population. Clones chosen after randomization, either initially ON (A, left) or OFF (B, left), will both randomize again upon induction in galactose (right). Expression is normalized to the mean of the pre-induction control lacking the recombinase plasmid. N = 10,000 cells per plot.

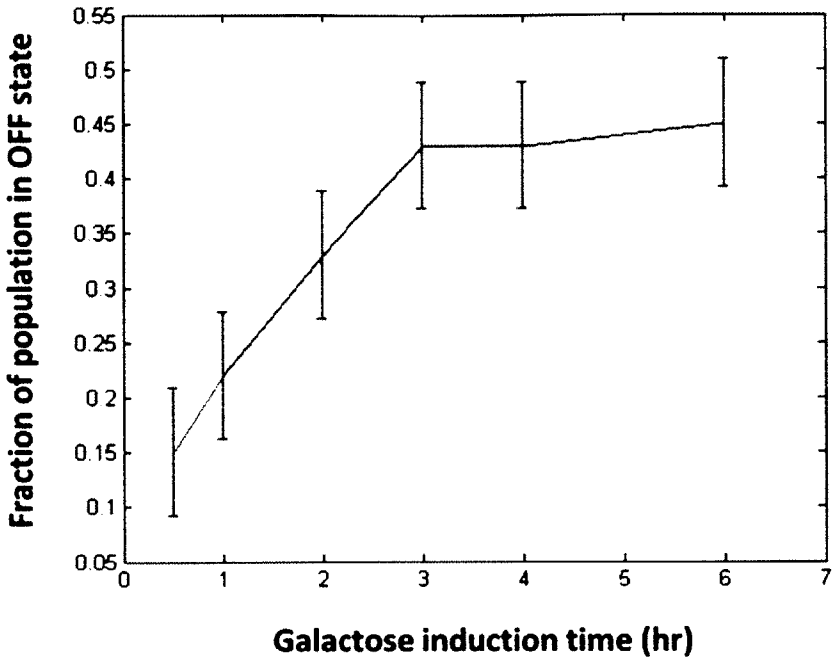


Figure 8: Kinetics of promoter inversion. A strain containing a floxed P_{TEF1}-YFP integrated at the *URA3* locus initially in the ON state, was induced and the fraction of the population in the OFF state was monitored for different times of galactose induction using fluorescence microscopy. Each distribution contains measurements of ~400 cells. Post-induction, cells are grown in synthetic media with glucose for 18 hours prior to microscopy. The 45%/55% OFF/ON split seen here is different from the 70%/30% split reported in the main text when cells were grown for 24 hours in glucose. We suspect some cells had not diluted out the YFP reporter at the 18 hr time point and were erroneously detected as ON, which is why we used 24 hours growth in glucose in all subsequent experiments. Nevertheless, regardless of galactose induction time, all samples were grown for 18 hrs in glucose. Hence we can conclude that after 4 hours of induction the fraction of OFF and ON cells does not change.

2.4 Independent randomization of multiple genes targets

We next sought to test the feasibility of randomizing multiple floxed promoters present in the same cell, as each additional promoter results in a combinatorial increase in the number of achievable genotypes. To do so, we constructed a haploid strain with 3 copies of the floxed *TEF1* promoter integrated at the *URA3, HIS3*, and *TRP1* loci, driving distinct fluorescent proteins. After induction of Cre in galactose, we found the 8 possible expression phenotypes within a randomized population present in a manner consistent with independent flipping of each cassette (Figure 9). We then mated haploid strains to create diploids with 4, 5, or 6 copies of the promoter. Five copies of the 2-state promoter randomized independently (Figure 10), but curiously repeated attempts of randomizing 6 copies failed, even with varying the time of Cre expression. Different attempts resulted in correlated expression states, with most of the genes either all ON or all OFF. Figure 11 shows increasing correlation among two of the reporters in a 6-copy strain. As Cre expression time increases, the side populations where one reporter is ON while another is OFF gradually disappear. One possibility for the failure of the 6 copy 2-state promoter may be due to the greater chance of recombination between loxP sites on different promoters, leading to large scale rearrangements. Therefore, with little or no recombination occurring, all the genes would initially be in the ON state, and when recombination occurs, rearrangements could lead to all genes in the OFF state.

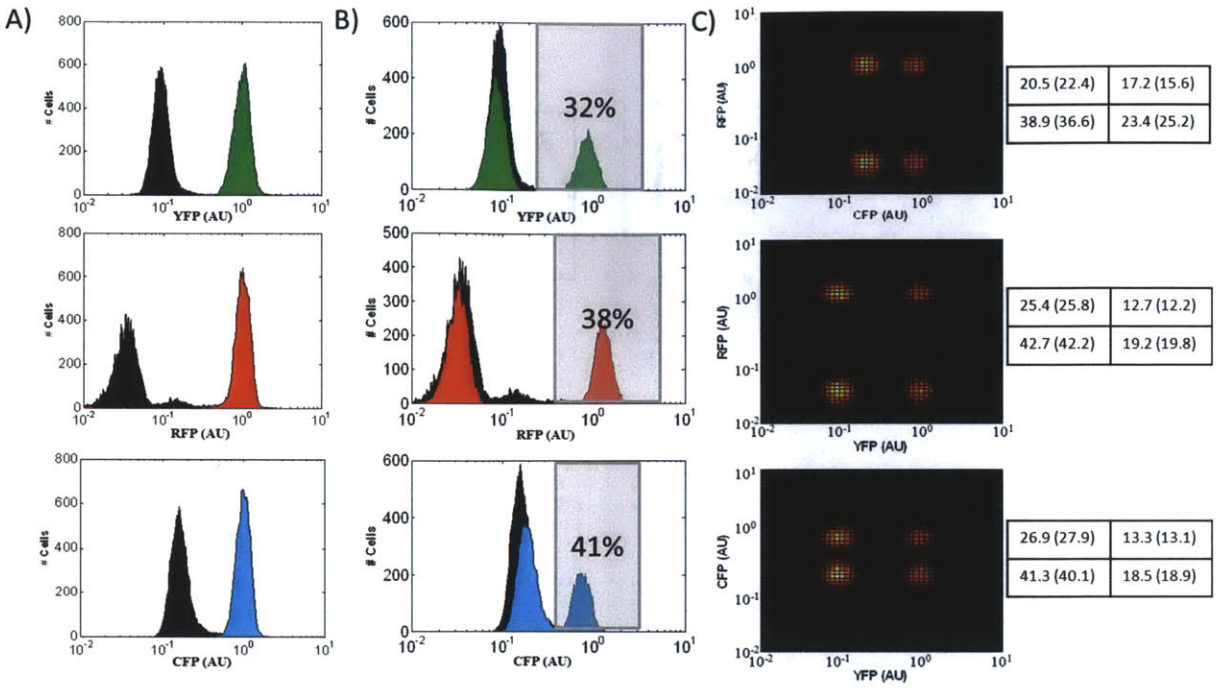
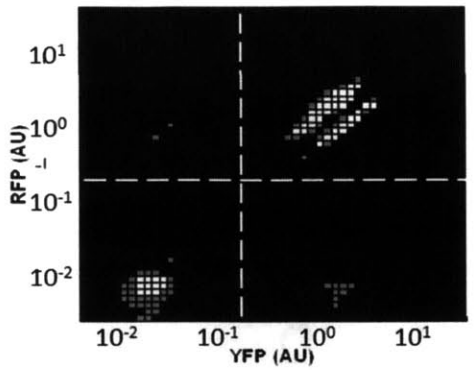
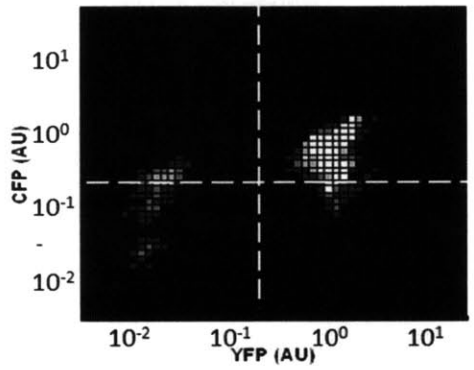


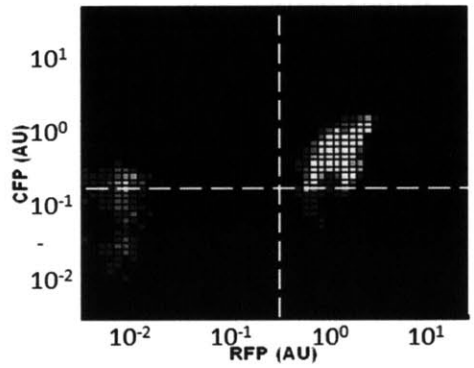
Figure 9: Three floxed promoters randomize orientation independently when present in the same cell. Flow cytometry data of a strain containing *URA3*::floxed P_{TEF1}-CFP, *HIS3*::floxed P_{TEF1}-RFP, and *TRP1*::floxed P_{TEF1}-YFP. Reporter fluorescence before randomization A) and after 12 hours of induction B) is shown normalized to the initial expression level. Autofluorescence of a strain lacking any fluorescene reporters is shown in black. The population fraction in the ON state is given for each individual reporter. C) 2D histograms of different pairs of the three reporters indicate independent randomization of each locus. The table to the right of each plot gives the population fraction in each quadrant, with the expected fraction in parentheses assuming completely independent randomization among reporters. N = 10,000 cells in each plot.



3.5 (15.9)	73.9 (61.5)
17.0 (4.7)	5.5 (17.9)



6.7 (16.2)	72.1 (62.6)
13.9 (4.3)	7.4 (16.9)



7.8 (17.7)	70.9 (61.0)
14.7 (4.8)	6.6 (16.5)

Figure 10: Randomization of a strain containing five floxed promoters. Comparative 2D histograms of reporter pairs in a diploid 5-color strain containing 2X (*URA3*::floxed P_{TEF1}-CFP), 2X (*HIS3*::floxed P_{TEF1}-RFP), and *TRP1*::floxed P_{TEF1}-YFP. Each plot is divided into quadrants according to ON versus OFF expression levels. The population fraction in each quadrant is reported to the right of the plot, with expected fractions in parentheses if reporters are randomized independently. N = 10,000 cells in each plot.

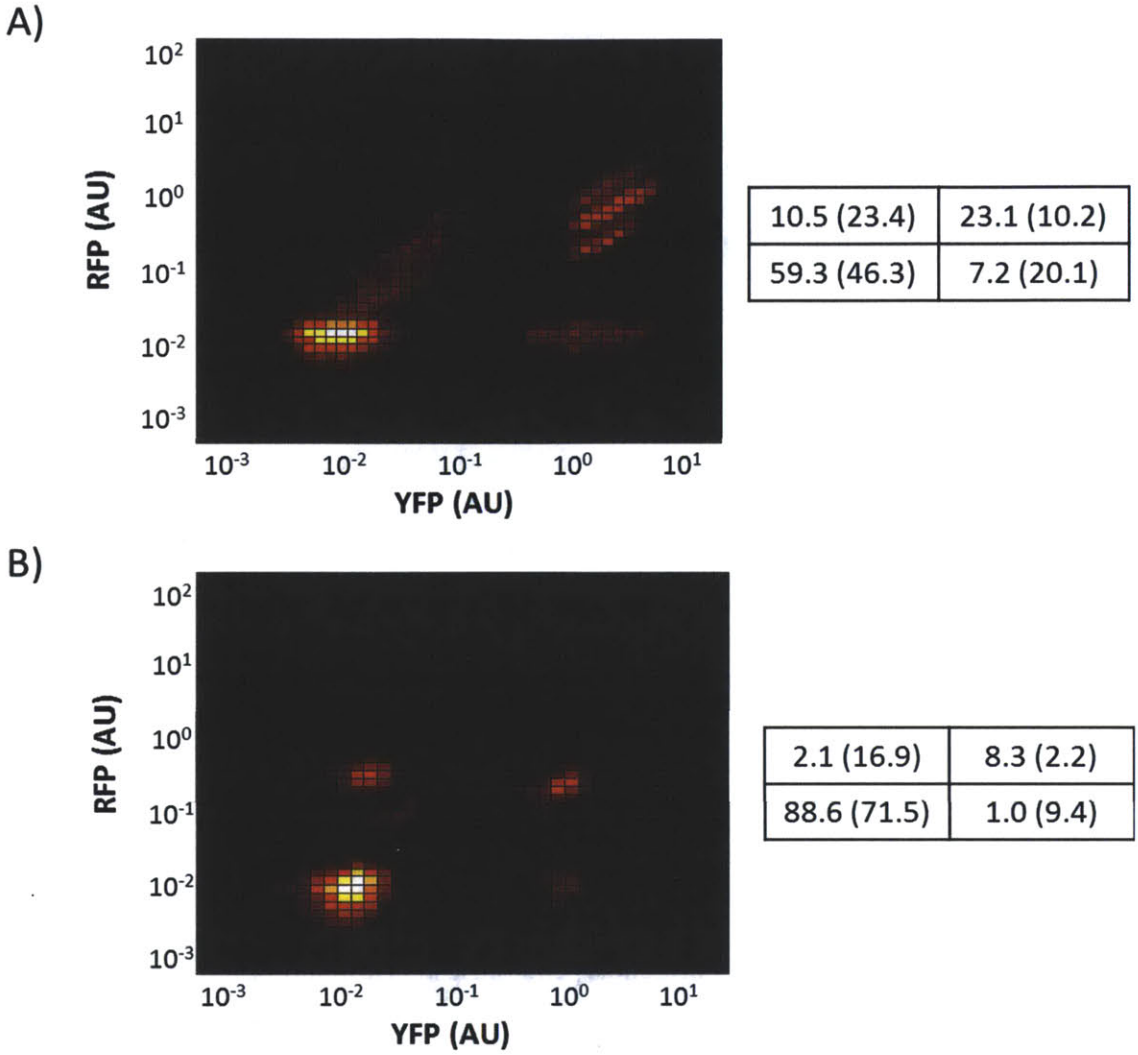


Figure 11: Reporter correlation in a strain containing 6 floxed promoters. 2D histograms of the YFP and RFP reporters with 6 hours (A) and 12 hours (B) of galactose induction in a diploid strain containing two copies each of *URA3*::floxed P_{TEF1}-CFP, *HIS3*::floxed P_{TEF1}-RFP, and *TRP1*::floxed P_{TEF1}-YFP. The population fractions in each quadrant are shown to the right, with the expected values in parenthesis assuming independent randomization. N = 10,000 cells per data set.

While small numbers of 2-state promoters recombine and randomize successfully, increasing the number of floxed promoters leads to increased correlation and disappearance of mixed populations. Further addition of floxed promoters would exacerbate this problem. To address this issue, we utilized orthogonal loxP site variants loxN and lox2272. We verified the invertability

of loxN- and lox2272- flanked promoters (data not shown), and then added 2 copies of a loxN-flanked *TEF1* promoter to make a new 6 copy strain.

To randomize this strain we found it best to express Cre for 12 hours (Figure 12). We then chose and analyzed the expression of 187 clones using flow cytometry. However, solely assaying expression was inadequate in completely characterizing the genotype because we were unable to distinguish clones that have one versus two promoters in the ON state driving expression of any particular reporter (Figure 13). While the range of YFP and RFP ON expression in Figure 13 was larger than that normally observed for a single reporter, a PCR scheme similar to that in Figure 6 did not yield a definitive cut off point between one or two ON reporters of a particular color. We assayed for promoter orientations in 38 of the clones to compare reporter expression and genotype. Because CFP and YFP share nearly identical sequence homology, the PCR strategy was also limited in uniquely determining the genotype. Still, we compared the combined results of the fluorescence and PCR assays in Table 3, in which each entry represents either a unique genotype or a small combination of genotypes. Entries corresponding to genotypes that should never be realized are shaded gray. We found the number of clones observed with a particular fluorescence phenotype and PCR-based genotype was largely consistent with what would be expected assuming each allele behaved independently. In addition, we confirmed at least partial independent randomization of the RFP promoter, since the P_{TEF1}-RFP locus can be uniquely assayed via PCR. In each case where the RFP reporter is ON, we found some clones which only give an ON band for RFP, and some which give both ON and OFF bands. One clone found in the gray region lacked only YFP expression which was confirmed multiple times. However, we could not verify the floxed promoter was in the OFF position by PCR. Most likely there was an irreversible loss of the floxed promoter upstream of YFP.

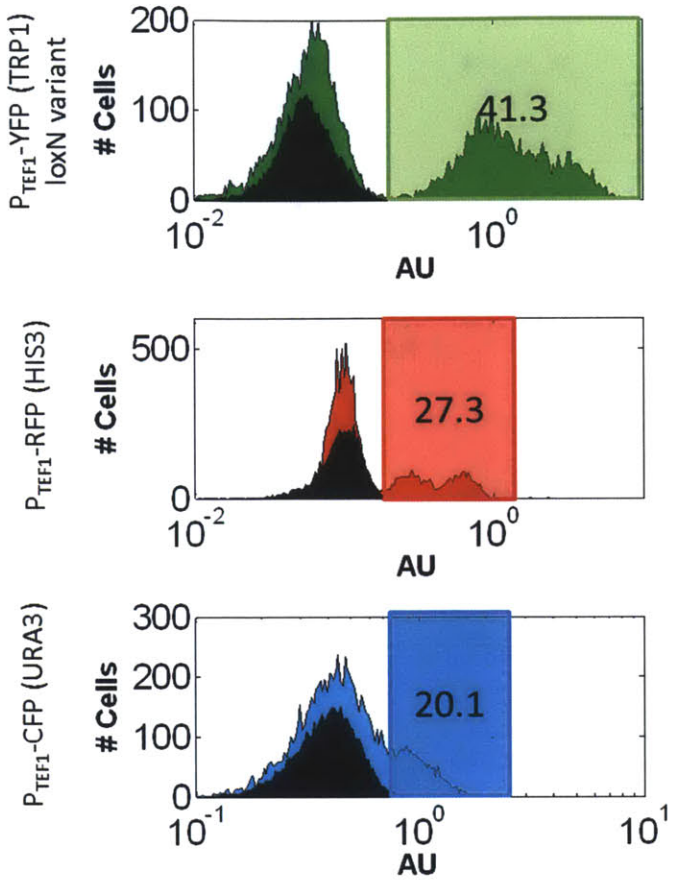


Figure 12: Single-channel reporter fluorescence histograms of a strain containing 6 floxed promoters. Randomization of expression in a diploid containing two copies each of *URA3*::floxed P_{TEF1}-CFP, *HIS3*::floxed P_{TEF1}-RFP, and *TRP1*::floxed P_{TEF1}-YFP. Autofluorescence is shown in black and the colored boxes show the range considered to be 'ON' and the percent of the population expressing in that range for each reporter. Expression is normalized to the mean expression of the un-randomized strain. N = 10,000 cells.

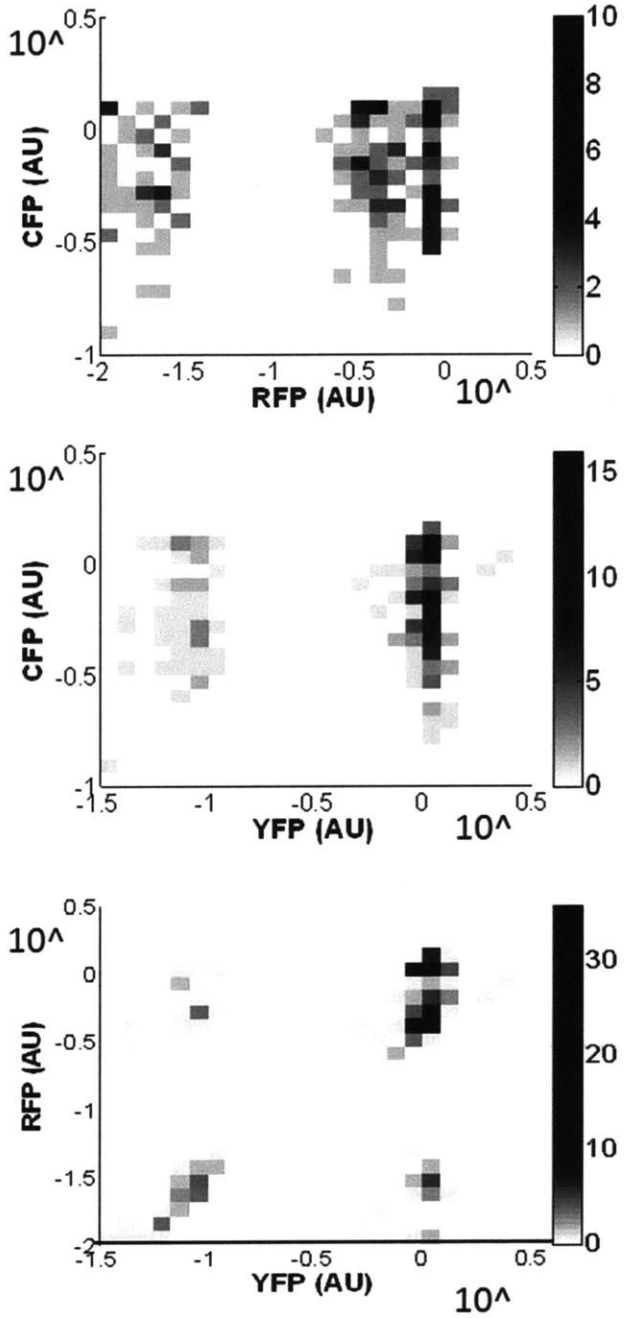


Figure 13: 2D histograms of fluorescence means from clones selected after randomization of the 6-color strain. Flow cytometry was used to measure the fluorescence profile of each of 187 clones picked from a randomized population. N = 10,000 cells per clone.

Table 3: Six copy 2-state promoters produce a wide range of genotypes after switching.

Mol.State										
Expr State	RFP	ON	ON/OFF	OFF	ON	ON/OFF	OFF	ON	ON/OFF	OFF	**Total**[b]
RFP/YFP/CFP	XFP	ON	ON	ON	ON/OFF	ON/OFF	ON/OFF	OFF	OFF	OFF	
ON/ON/ON		5 (1.8)[a]	3 (1.4)		0 (3.8)	2 (3.1)					10/37
ON/ON/OFF					0 (2.2)	2 (2.8)					2/71
ON/OFF/ON			1 (0)		2 (2.8)	2 (2.2)					5/10
ON/OFF/OFF								4 (3.3)	2 (2.7)		6/15
OFF/ON/ON				2 (0.9)			2 (2.1)				4/10
OFF/ON/OFF							3 (5)				3/12
OFF/OFF/ON							4 (4)				4/11
OFF/OFF/OFF										4 (4)	4/21

[a] Value is the number of cells found in the particular molecular state (as determined by PCR). Parenthetical values indicate expected number of cells in each molecular state assuming each allele inverts independently, given the observed fluorescent expression state.

[b] Number of single clones with the observed fluorescent expression state (167 total). 38/167 were probed by PCR

2.5 Randomization of a '4-state' promoter

We also extended the recombinase strategy to create a 4-state cassette that leads to four expression states (OFF, LO, MED, or HI) that span a > 100-fold range. The 4-state cassette consists of two pairs of floxed promoters in a divergent orientation. In the presence of Cre, one pair of promoters may be irreversibly excised, but the remaining promoter pair is still capable of inversion events (Figure 14). We first tested the functionality of one copy of the 4-state cassette integrated at the *URA3* locus. Upon randomization, we observed the anticipated range of reporter expression levels in the population. Using primers specific to each promoter, we confirmed that recombination resulted in any of the four promoters driving reporter expression and that one of the promoter pairs is excised during randomization. Individual clones chosen from the population exhibited distinct expression levels indicative of one of the four promoters (Figure 15).

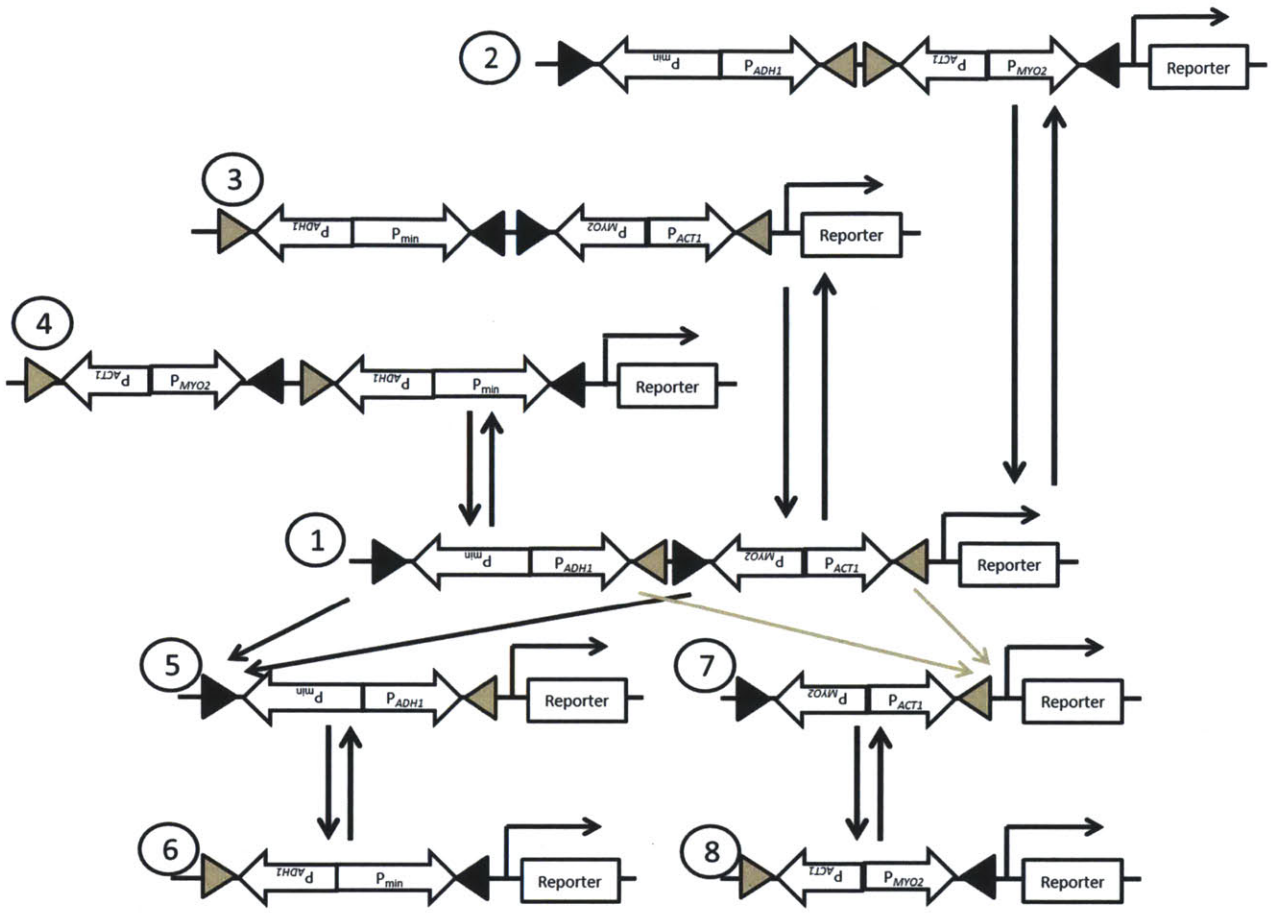


Figure 14: Sample recombination events of the four-state promoter. The initial orientation of the 4-state promoter (1) contains two promoter pairs, each flanked by lox sites, and may undergo reversible recombination involving only one of the promoter pairs (2 or 3), or inversion of the entire cassette (4). Each of these products may undergo further reversible recombination events whose products are not shown. Irreversible excision of one of the promoter pairs (5-8) is always observed in clones picked after expression of Cre, however the remaining promoter pair may still reversibly recombine (e.g. 5 and 6). Any of the intermediates 1-4 may result in the truncated products 5-8 after irreversible excision of one promoter pair. Different colored lox sites are meant as an aid to the reader and do not represent functional or sequence differences. Recombination between lox sites of the same color while in orientation 1 will result in either product 5 or 7.

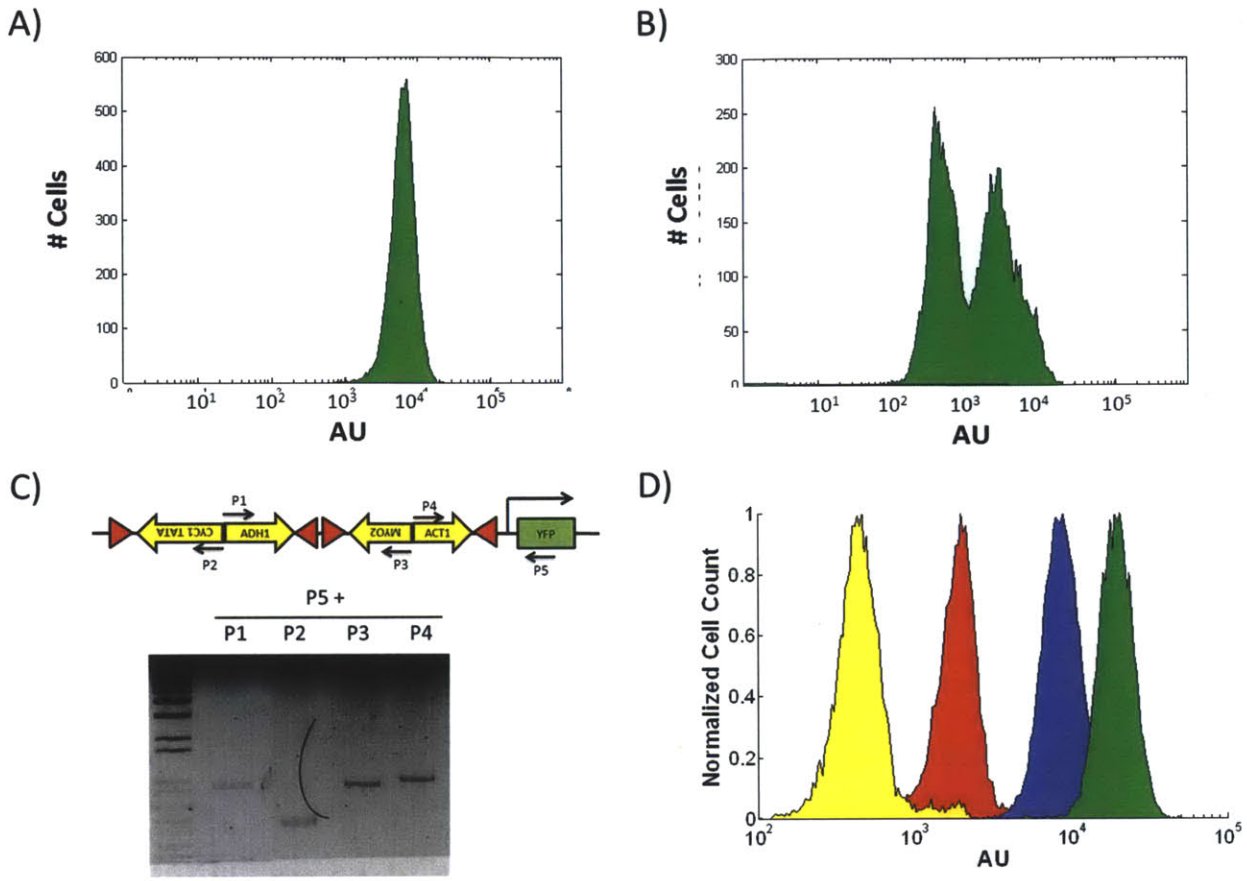


Figure 15: Randomization of the four-state promoter driving YFP at the *URA3* locus. Induction of Cre in an initially homogeneously expressing strain (A) generates a range of expression levels (B). Bulk population PCR post-randomization shows that recombination events result in all four promoters driving the reporter (C). D) Individual clones selected after randomization and verified with PCR show that each promoter has discrete expression levels over the range initially seen in the bulk population. Promoters (and corresponding colors) in (D) are minimal TATA box (yellow), *MYO2* (red), *ADH1* (blue), and *ACT1* (green). N = 10,000 cells per data set.

We then mated two haploid strains with the integrated 4-state cassette driving YFP and RFP, both integrated at the respective *URA3* loci, to generate a two copy 4-state diploid. Upon randomization, we observed a similar range of expression levels with little correlation in the expression state of each reporter (Figure 16B). We chose 60 individual clones and analyzed their expression profiles by flow cytometry. A subset is shown in Figure 16C. We confirmed the existence of all possible promoter rearrangements in the randomized population with a PCR scheme similar to **Figure 15**C, but with unique products for YFP and RFP (Figure 16D), and

used the same scheme to identify specific genotypes among the chosen clones (Figure 16E). Figure 16B includes expression ranges of each of the 4 promoters (top and right of the plot) taken from the analyzed clones.

While loss of one promoter pair is expected for the 4-state cassette in the presence of Cre, we expected to find one intact promoter pair upstream of the fluorescent reporters in each clone. However, in some cells with no reporter expression we were unable to detect the presence of any part of the cassette by PCR (Figure 16E, far right). We suspected in these cases an irreversible recombination event may have deleted the cassette, as they also had no reporter expression for the reporter/promoter pair which could not be detected. To test this possibility, we subjected all of the clones to another round of galactose induction for six hours. Approximately 25% of the loci in the clones failed to re-randomize (15/60 for YFP, 17/60 for RFP). For any application that does not require repeated randomization, this irreversibility should still be acceptable because all possible genotypes are realized within the population.

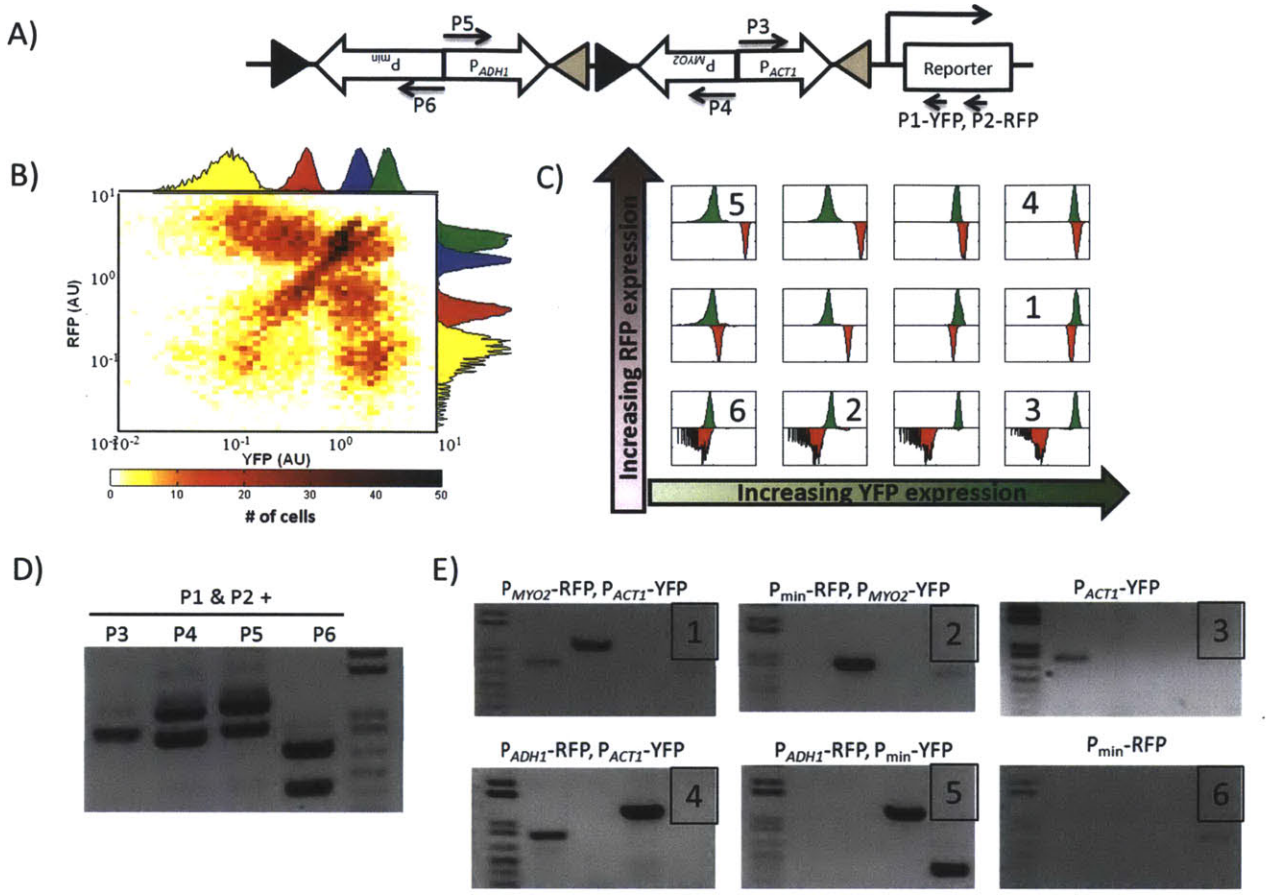


Figure 16: **Randomizing gene expression level using four-state promoters.** A) Placing two floxed pairs of native promoters in tandem creates a "four-state" promoter capable of randomizing the identity of the promoter proximal to the gene. Labeled arrows show a primer scheme to assay the identity of the promoter driving the reporter after randomization in the presence of Cre. B) 2D histogram of RFP and YFP expression measured from two different four-state promoters with six hours Cre induction. Fluorescence values of single cells are normalized to the average fluorescence of a non-randomized control strain with the strongest promoter P_{ACT1} driving expression (N = 9,449). 1D expression histograms on the edges of the 2D histogram show the tight population variability of each of the four component promoters: green-P_{ACT1}, blue-P_{ADH1}, red-P_{MYO2}, yellow-minimal TATA box. C) 1D expression histograms from clones picked from a randomized population. D) Genomic PCR from a randomized population using the scheme in (A). Larger bands in each column correspond to RFP, smaller ones to YFP. Four genotypes corresponding to each of the two tandem promoter pairs are seen in the population. E) Genotyping individual clones picked from a randomized population using genomic PCR as in D). In each gel, the four lanes following the ladder in each gel correspond to the primer combinations in D). Numbers on each gel label the clone and its corresponding fluorescence profile is also labeled in (C). Two of the clones (far right) are examples of strains in which one reporter cannot be detected through PCR and has an expression profile which is indistinguishable from autofluorescence measured from a strain containing no fluorescent reporter.

2.6 Discussion

By flanking promoters with inverted loxP sites, we have shown how to rapidly and repeatedly randomize the expression state of multiple promoters in budding yeast, leading to stable phenotypes which can be interrogated for desirable properties. This recombinase-based strategy may extend to the large number of organisms in which the Cre-loxP system has been shown to function, perhaps with modifications to prevent undesirable recombination events leading to chromosomal loss and cell death. We suggest it is particularly useful for interrogating or retooling branched pathways, whether part of metabolic, regulatory, or processing networks such as base excision repair. By choosing to randomize expression of proteins that dictate the activity or flux through various branches, the innate capabilities of the network to produce various outcomes or products can be explored. It also confers an additional advantage over traditional genetic methods that delete or overexpress genes. Because all target genes are initially expressed at native levels or overexpressed, certain genotypes may be present after randomization that would be difficult to construct with traditional methods. This would be true of genetic suppressors that consist of multiple genes -- for example, a combination of two sets of synthetic lethal genes that are not lethal when both sets are deleted in combination.

As the number of genomically integrated promoters and loxP sites increases, we find an increased probability of irreversibly ending in one expression state. Because we are unable to detect the state by PCR we surmise that Cre recombination results in some type of large scale deletion or rearrangement. One possibility is unequal exchange at sister chromatid leading to chromosome loss, as reported previously in mice. In haploid cells, the ensuing loss of essential genes would lead to cell death and may also explain the bias for OFF cells seen with a single floxed promoter copy. In diploid cells, though, loss of one chromosome could result in a viable

monosomy. This is a potential consideration against employing diploids, though it may not be a dominant event in yeast: multiple clones genotyped in the 6 color diploid were heterozygous for the RFP expression state (Table 3), which indicates rearrangements can occur without chromosomal loss. A second possibility is recombination between loxP sites in different floxed promoters, leading to some large scale but viable rearrangement. We found a critical parameter is the amount of time Cre is expressed, and tuning this time allowed randomizing up to 5 2-state promoters within one cell. Perhaps the high rate of chromosomal loss in mice due to inverted loxP sites is due to the level and/or constitutive duration of Cre expression. In order to make sure that all cells in the population express Cre long enough to facilitate multiple inversion events, but to decrease the chance of accumulation of large-scale rearrangements or deletions, we recommend expressing Cre for 6-12 hours.

In principle, utilizing the other two loxP variants could allow 15 2-state promoters to be randomized. An alternative strategy to further prevent 'cross-talk' between multiple promoters is to express multiple versions of Cre with altered loxP site preferences or even other recombinases (Groth and Calos 2004). A similar strategy would apply to the 4-state promoter where using the same type of loxP site limited its use to 2 copies because of irreversible recombination (Figure 16). Still, by just employing the two other loxP variants potentially ~6 4-state cassettes could be integrated and randomized successfully, sampling $4^6 = 4096$ expression states with 2 sets of 3 integration steps followed by a single mating. These could be useful in cellular engineering contexts where fine-tuning of gene expression is the ultimate goal. However, as more genes are randomized simultaneously, keeping floxed promoters with identical loxP sites on different chromosomes should minimize irreversible inter-cassette recombination events. Placing the

floxed promoters on opposite arms of one chromosome may be sufficient although we have not tested this.

2.7 References

Adams, Bruce G. 1972. "Induction of Galactokinase in Saccharomyces Cerevisiae: Kinetics of Induction and Glucose Effects." *Journal of Bacteriology* 111 (2): 308-315.

Grégoire, Damien and Marie Kmita. 2008. "Recombination between Inverted loxP Sites is Cytotoxic for Proliferating Cells and Provides a Simple Tool for Conditional Cell Ablation." *Proceedings of the National Academy of Sciences* 105 (38): 14492-14496.

Groth, Amy C. and Michele P. Calos. 2004. "Phage Integrases: Biology and Applications." *Journal of Molecular Biology* 335 (3): 667-678.

Lee, Gwang and Izumu Saito. 1998. "Role of Nucleotide Sequences of< i> loxP Spacer Region in Cre-Mediated Recombination." *Gene* 216 (1): 55-65.

Lewandoski, Mark and Gail R. Martin. 1997. "Cre–mediated Chromosome Loss in Mice." *Nature Genetics* 17 (2): 223-225.

Longtine, Mark S., Amos McKenzie III, Douglas J. Demarini, Nirav G. Shah, Achim Wach, Arndt Brachat, Peter Philippsen, and John R. Pringle. 1998. "Additional Modules for Versatile and Economical PCR-Based Gene Deletion and Modification in Saccharomyces Cerevisiae." *Yeast* 14 (10): 953-961.

Nagy, Andras. 2000. "Cre Recombinase: The Universal Reagent for Genome Tailoring." *Genesis* 26 (2): 99-109.

Sauer, Brian. 1987. "Functional Expression of the Cre-Lox Site-Specific Recombination System in the Yeast Saccharomyces Cerevisiae." *Molecular and Cellular Biology* 7 (6): 2087-2096.

Sikorski, Robert S. and Philip Hieter. 1989. "A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in Saccharomyces Cerevisiae." *Genetics* 122 (1): 19-27.

Chapter 3: Modulating the fates of DNA lesions by randomizing expression of various repair enzymes

3.1 Abasic site repair as a target system for promoter recombination

In Chapter 2 we showed that randomizing promoter orientation resulted in the desired genotypic diversity at target loci by monitoring fluorescent reporter expression. To further demonstrate the application of the DNA recombinase strategy we sought to randomize the expression levels of a group of endogenous genes involved in repair of apurinic/apyrimidic (AP) sites, also called abasic sites, and select for individual variants that could tolerate high levels of DNA damage.

Extra- and intra-cellular factors constantly damage DNA molecules and bases. Among other forms of damage, small, non-helix distorting DNA lesions such as alkylation damage are recognized by DNA glycosylases, which remove the damaged nucleotide and leave an abasic site. In most cases, repair pathways identify and fix the site before replication in an error-free way; however if the site persists until DNA synthesis occurs it can cause replication fork stalling and collapse. Error-prone polymerases, which lack much of the proofreading capability of high fidelity DNA polymerases, are able to bypass damaged sequences and allow the replication fork to continue, but at the cost of much higher rates of mutation at the damaged locus (Prakash, Johnson, and Prakash 2005; Boiteux and Guillet 2004; Memisoglu and Samson 2000).

In yeast the genes *APN1* and *APN2* encode AP endonucleases that cleave the phosphodiester linkage at the AP site, initiating further events that lead to error-free base excision repair (BER) (Demple and Harrison 1994; Johnson et al. 1998). *RAD4* can also identify and mark AP sites for repair via recruitment of components of the nucleotide excision repair (NER) pathway, a more general error-free repair pathway capable of targeting AP sites as well as larger DNA lesions

(Xiao and Chow 1998). If unrepaired AP sites cause the replication fork to stall during DNA synthesis, repair can proceed through an error-free or a mutagenic pathway. The mutagenic lesion bypass pathway requires an "inserter" polymerase to insert a base opposite the AP site, and the extender polymerase, Pol ζ, a heterodimer of the gene products of *REV3* and *REV7* (Haracska et al. 2001). Thus, stalled replication forks can be restarted, sometimes at the cost of mutations (Figure 17).

Although *APN1* constitutes the major source of AP endonuclease activity, deleting either *APN1* or *APN2* has no phenotype (Johnson et al. 1998; Ramotar et al. 1991). The *apn1* background does have a mild sensitivity to methyl methane sulfonate (MMS), a DNA alkylating agent which indirectly increases the AP site load, resulting in increased mutation rates. The *apn1 apn2* background exhibits a much more pronounced MMS sensitivity, with a much larger increase in mutation rate. Deletion of either *REV3* or *REV7* in the *apn1 apn2* background reduces the elevated mutation rate seen both in the presence of absence of MMS, confirming the importance of Pol ζ in mutagenic lesion bypass (Johnson et al. 1998).

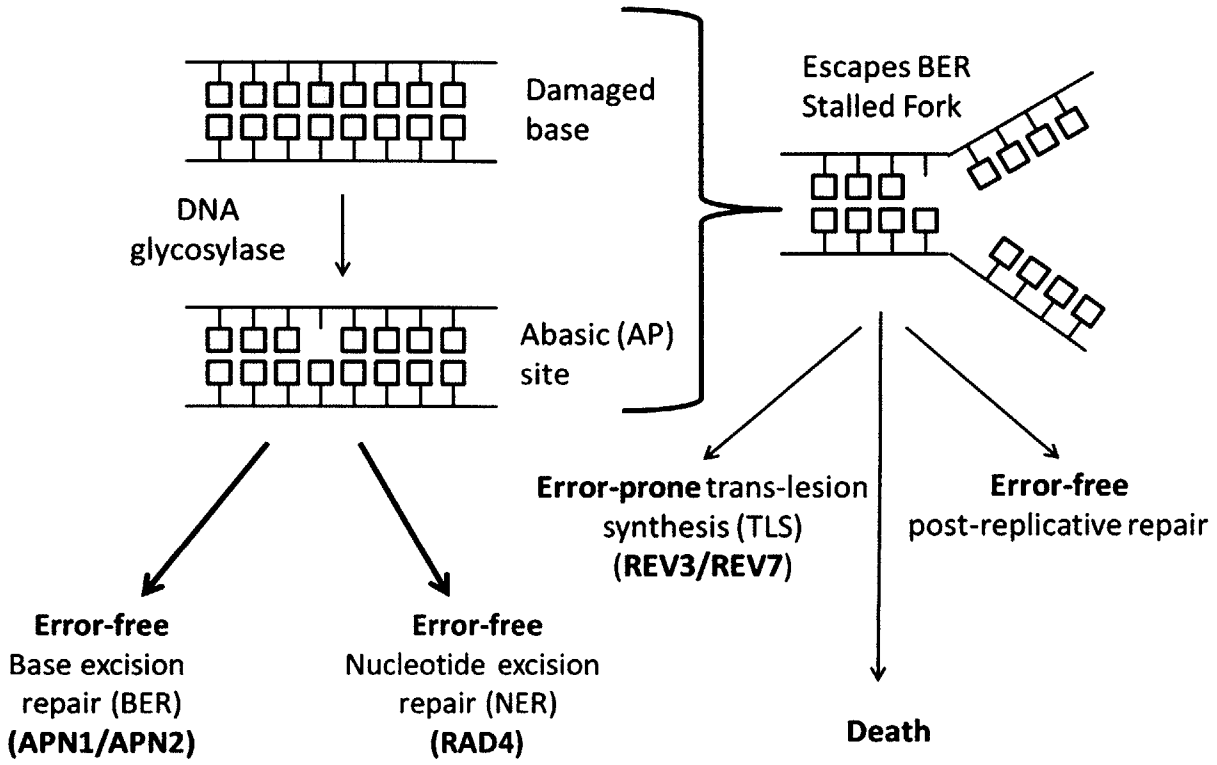


Figure 17: Role of selected target genes in repair of DNA lesions. DNA bases with damage from chemical modifications are first converted to abasic sites. The error-free repair of these sites begins with cleavage of the phosphodiester linkage by AP endonucleases, coded by APN1 and APN2 in yeast. Unprocessed abasic sites can lead to replication fork stalling during DNA synthesis. The replication fork is potentially restarted by error-prone trans-lesion polymerases, including the heterodimeric Pol ζ, which is coded by the REV3 and REV7 genes in yeast.

We reasoned that modulating the expressions states of *APN1, APN2, REV3, REV7,* and *RAD4* would change the repair capacity of AP sites through various pathways (Figure 17) and ultimately affect the growth rate, especially in the presence of MMS, because of the cell cycle-dependent nature of the modes of repair as well as their respective rates. Increasing the expression of each gene should increase repair capacity of the cell, but it is difficult to anticipate how increasing or decreasing the load of repair flux will contribute to survivability and growth rate. Increasing the expression of BER and NER genes should decrease the amount of abasic sites which persist into replication, while increasing the expression of error-prone polymerases

may decrease the chance that accumulated damage will lead to checkpoint arrest. Conversely, increasing replication-dependent repair may not increase survivability if the load of abasic sites during replication is too great, even with elevated Pol ζ expression, and may also decrease growth rate as the repair methods during replication are time consuming.

Previous studies indicate the *apn1 apn2* and *apn1 apn2 rev3* and *apn1 apn2 rev7* backgrounds exhibit reduced growth rates in the presence of MMS (Johnson et al. 1998). However, there are limited studies looking at the effects of overexpressing some of these genes in the presence of a high AP site load, particularly when other genes are compromised. Both *REV3* and *REV7* have been moderately overexpressed (1.8 and 3-fold) to examine the effect of increased Pol ζ activity in the presence of UV. This lead to a mild (3-fold) increase in the background mutation rate but increased UV sensitivity and a decreased growth rate (Rajpal, Wu, and Wang 2000). This result hints that increasing repair through mutagenic pathways may always result in decreased growth fitness and that overexpression of all five genes may not be obviously beneficial in DNA damage stress environments.

We randomized expression of these genes between an OFF state similar to a deletion, an ON state where the gene is constitutively expressed, and a NATIVE state representing native regulation. We subjected the randomized population to a high AP site load by addition of MMS to identify genotypes resistant to high levels of DNA damage (Section 3.3). We found that increased expression of BER enzymes imparted higher survivability, with a possible contributing factor from *REV3,* but that overexpression of Pol ζ played no role in this phenotype. We also found that randomization of the genes (both with and without *RAD4*) led to increased variability in growth fitness in rich media as well as DNA damage stress conditions.

3.2 Combining a strain building technique and promoter recombination to randomize gene expression among three discrete states

We sought to randomize the expression of the *APN1*, *APN2*, *REV3*, and *REV7* genes using the 2-state promoter. However, the native regulation of an endogeneous gene represents an important additional expression state. To create a library with (3 expression states)$^{4\ genes}$ = 81 phenotypes, we began by constructing two haploid cells of opposite mating types as follows. In each haploid, promoters of 2 of the 4 genes were replaced with a floxed promoter using the disruption vectors detailed in Figure 4B. The haploids were then mated and sporulated. Because only particular haploids and unsporulated diploids had canavanine sensitivity, we could specifically select for haploids after sporulation (Figure 18).

We observed diversity in growth rates by monitoring colony size distributions of randomized strains compared to isogenic controls. We verified the presence of all 3 expressions states for each gene in the bulk mixed population after induction of Cre using PCR (Figure 19). All four floxed promoters are present on different chromosomes, minimizing the possibility of interpromoter recombination. We plated the library on rich media (YPD) and rich media containing 0.0007% MMS and compared colony sizes after 24 hours of growth (Figure 20A). The two haploids used for mating differed at the *ADE2* locus (*ADE2+* and *ade2-1*) and adenine autotrophy was not selected after sporulation, so as controls we plated haploids isogenic to both but without floxed promoters. While the loss of the *ADE2* auxotrophic marker had a significant effect on decreasing colony size, the mixed populations both before and after randomization exhibited a range of colony sizes that spanned the different ranges seen in both controls. For rich media, the large population fraction with low growth rates suggested at least some of the difference was due to differences in DNA repair gene expression and not due to differences in

the *ADE2* locus. We also observed a pronounced difference in colony size distributions in the case of low levels of MMS. Control strains exhibited a decrease in the mean and range of colony sizes upon MMS addition. Both the non-randomized and randomized library did not experience this decrease, and the colony size distribution of the randomized library retained a larger variance. These results also suggest that randomization of the target genes affected phenotypes within the population, with the possibility that certain genotypes confer an advantage to growth in MMS compared to the control strains. Replicates of strains plated at different densities show that colony size measurements are reproducible with a sample size larger than 100 colonies per plate (Figure 20B) with the median fluctuating only 5% among replicates and with little change in the span of the population distribution.

The minimum colony size observed in Figure 20A indicates a lower limit for comparing colony size distributions. This limitation is due to our inability to observe differences in size when colonies are small. Size was scored computationally with a minimum pixel size set to correctly identify colonies, meaning that populations which contain colonies in this size range will appear to have identical lower limits. However, if the colonies are allowed to grow for longer periods of time to try and detect small size differences, the differences between colony size distributions as a whole become less pronounced. This is because of a growth-rate independent plateau in colony size that is probably due to nutrient limitations.

Despite these limitations, we can also infer the growth rate in liquid media from colony size data. Figure 21 shows a linear correlation between log colony area and liquid growth rate where a 2-fold increase in log colony area corresponds to ~3.5-fold increase in liquid growth rate. Using colony area as a measure of growth rate could allow more thorough screening of the library as hundreds of colonies from a randomized population can be plated and imaged within one day

compared to measuring the liquid growth rates of hundreds of clones in triplicate, provided the colony size distributions of single clones are tightly distributed and hence distinguishable.

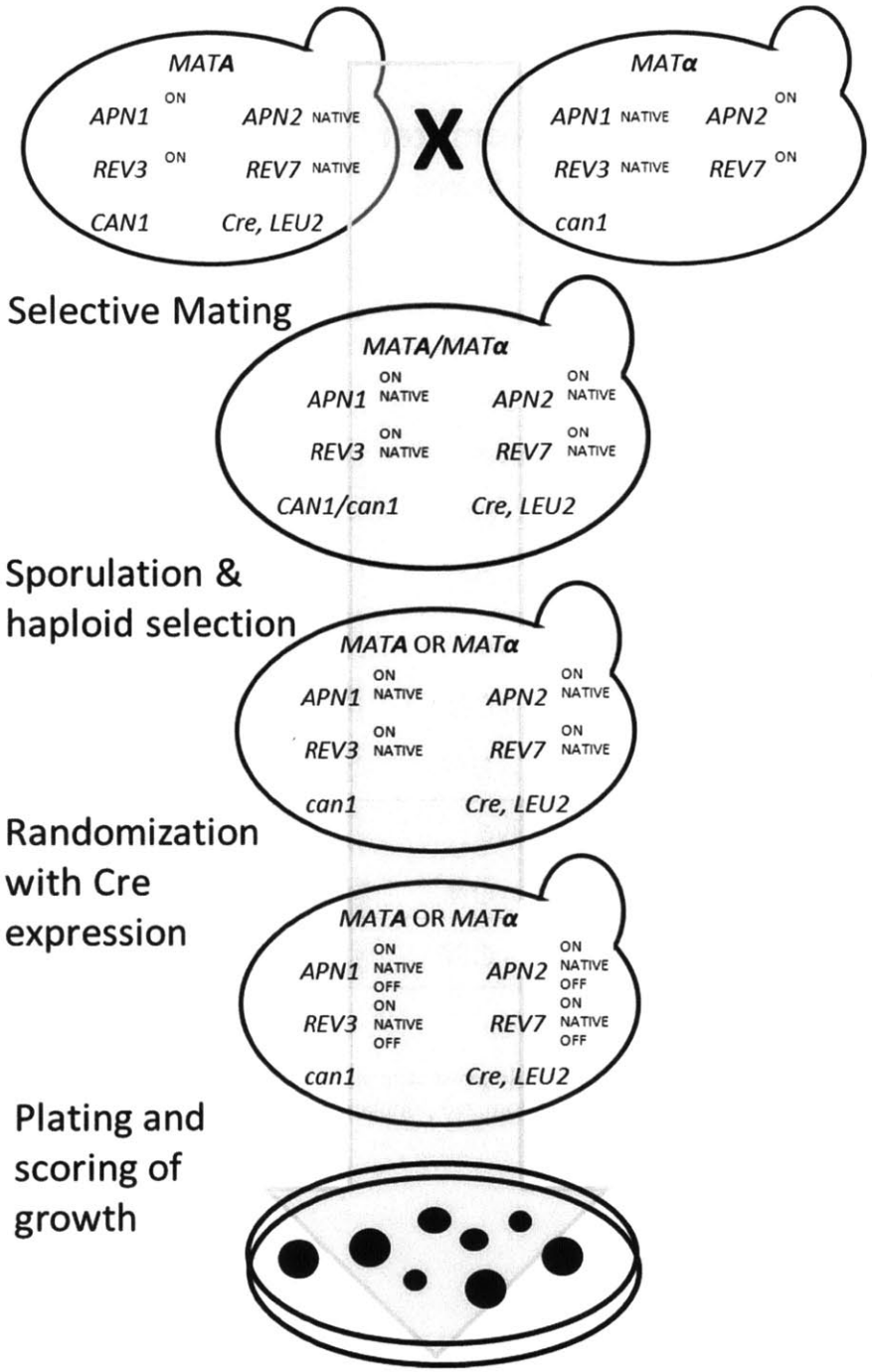


Figure 18: A protocol for generating a gene expression library containing combinations of three expression levels at desired loci. At each step, the expression levels which exist in the population for each gene are written to the right of the gene name. In the first two steps, the populations are isogenic, with each cell containing the

expression states shown next to each gene. After sporulation, random assortment results in a genetically heterogeneous haploid population. Finally, randomization with Cre expression results in a population where each target locus has one of three possible expression levels. These populations can then be plated with or without selection and phenotypically scored based on colony size.

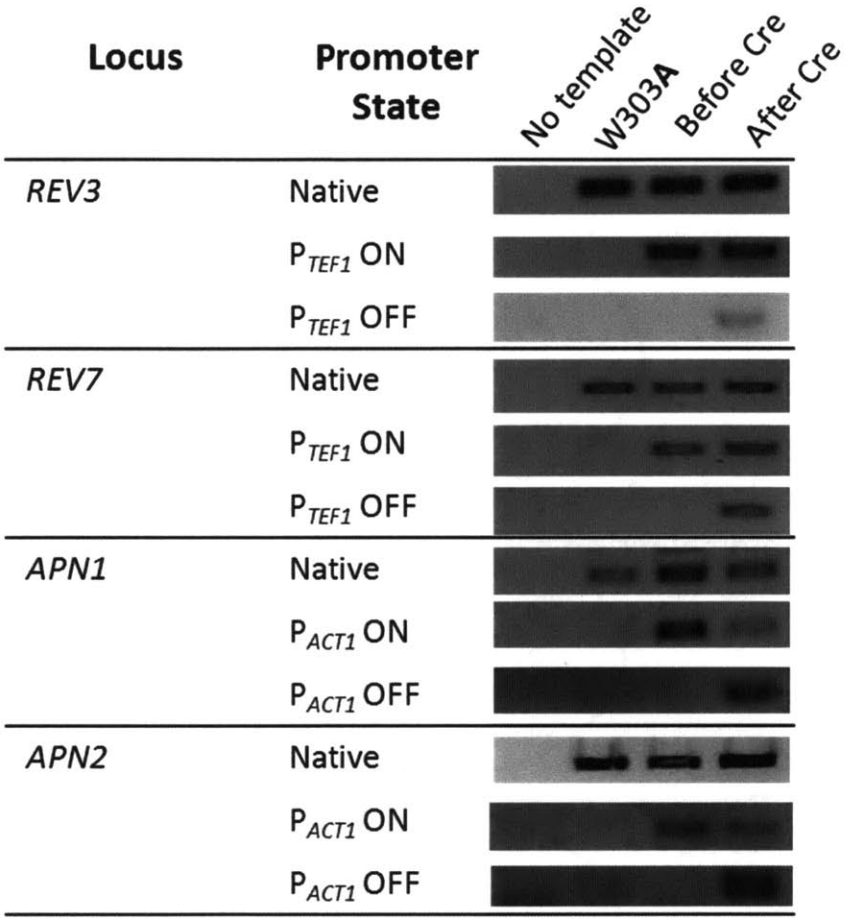


Figure 19: Recombination generates all possible promoter states in a randomizaed population. PCR using genomic DNA isolated from the sporulated, randomized population indicates all possible genotypes are present at each locus.

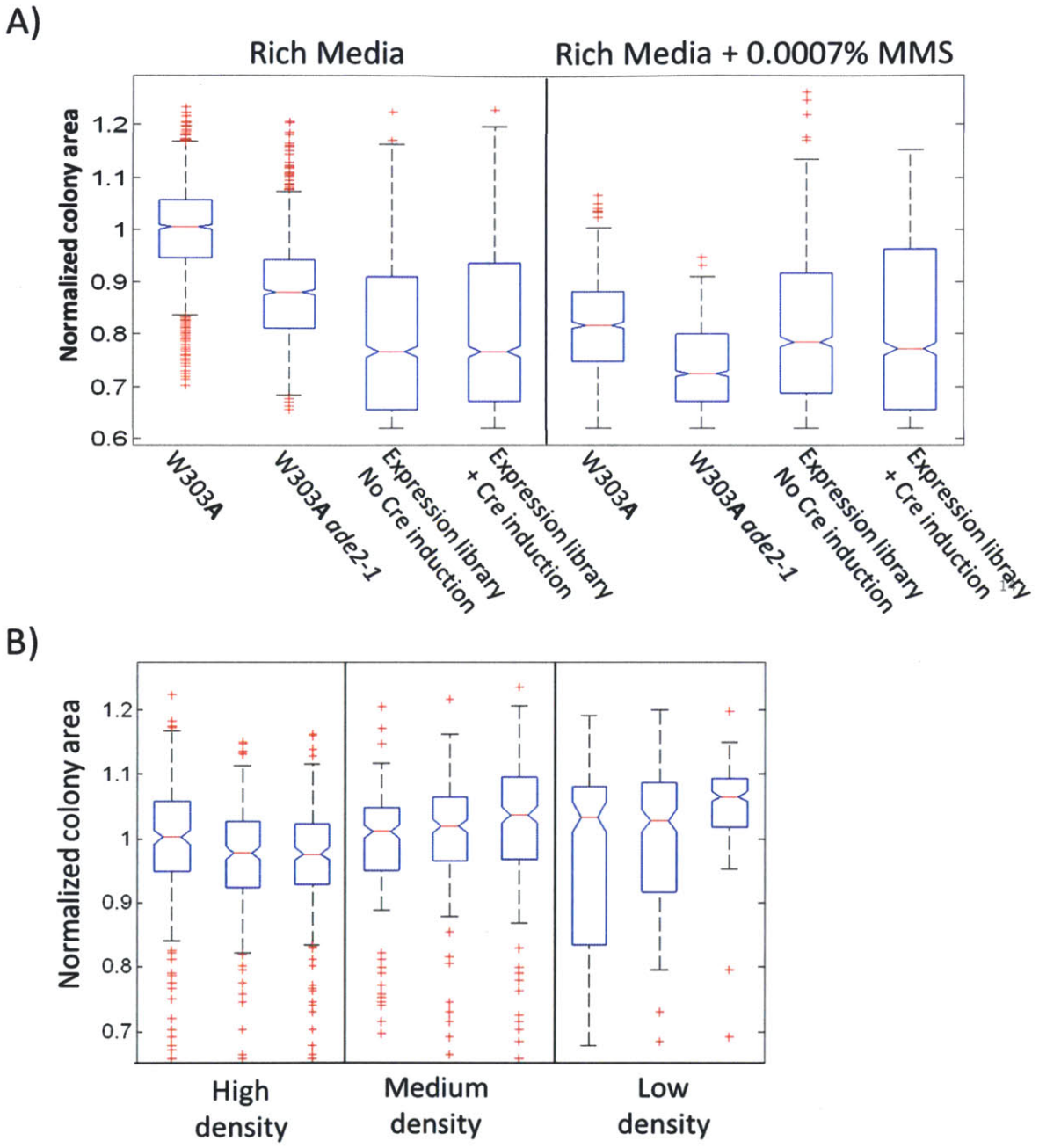


Figure 20: Randomization of DNA repair enzymes generates growth phenotype diversity. A) Boxplots of colony size distribution on YPD plates after 1 day of growth without MMS (left) and with 0.0007% MMS (right). Each population was grown in liquid YPD for 24 hours before plating. The "Normalized Colony Area" refers to the ratio of the logs of individual colony areas by the median of the haploid wild type population with intact *ADE2*. The box encompasses the middle 2 quartiles of the distribution and the whiskers extend to one box length or to the farthest outlier if it is within one box length. Crosses are individual points outside the whisker range. . N = 1319, 1506, 1119, 758, 356, 153, 362, and 344 colonies, respectively. B) Boxplots of replicate plates of the control strain used in (A) plated at different colony densities. Each plot is normalized to the median of the left-most distribution. N = 245, 229, 243 (high density), 147, 162, 142 (medium density), and 47, 57, 38 (low density) colonies, respectively.

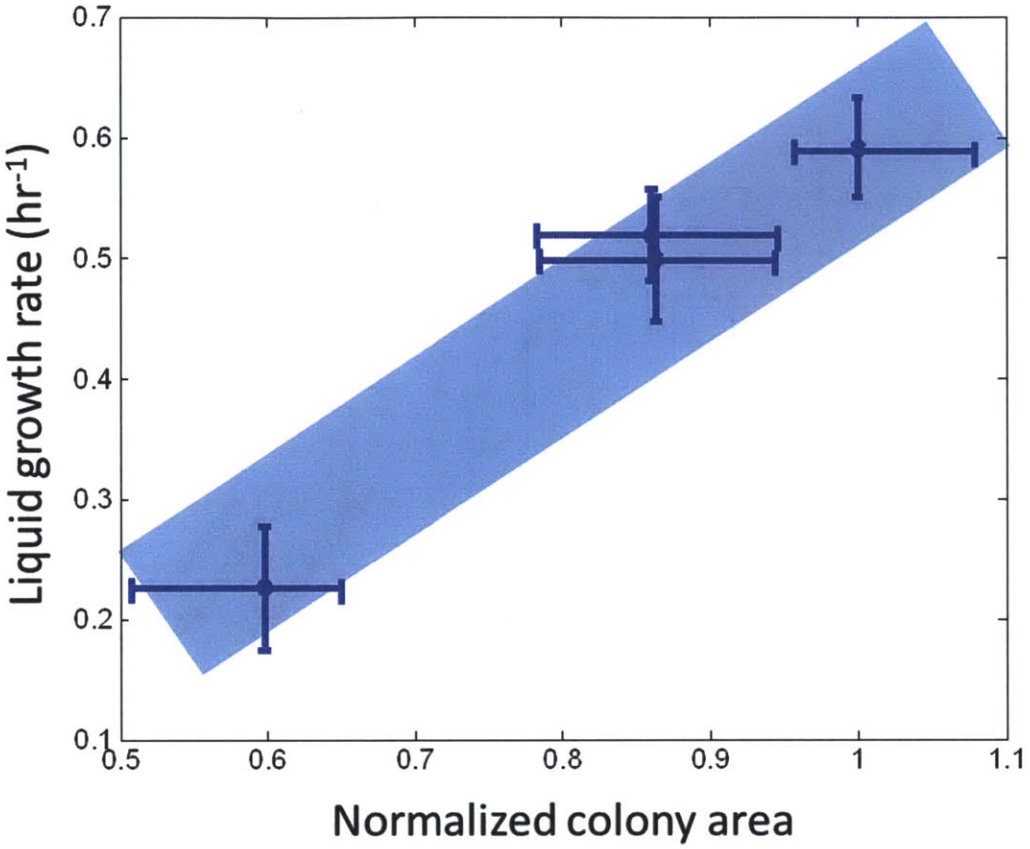


Figure 21: Liquid growth rate correlates proportionally with log colony area. Colony sizes after 1 day of growth on galactose of strains with different growth rates compared to growth in liquid YPD media. Vertical error bars are the standard deviation of the growth rate fitted to six replicates per strain. Horizontal error bars encompass the 2nd and 3rd quartiles of the log colony area of each strain with N = 1310, 1420, 1312, and 483 colonies from left to right, respectively. The linear bar is meant as a guide to the eye.

3.3 Increased expression of BER enzymes increases survivability in high MMS conditions

In addition to testing the ability of promoter recombination to affect growth fitness, we also subjected the DNA repair expression library to high levels of DNA damage stress to find a genotype which might increase the capacity for repair and growth in conditions where the widtype could not survive. We selected for clones that could grow in the presence of high levels of MMS by plating 400 cells (~5x coverage of 81 different genotypes in the library) on rich media (YPD) with 0.035% MMS. Sixteen colonies were observed after growth for five days compared to two colonies for a control with a wild type haploid plated. The growth rates of eleven clones in liquid YPD media + 0.0035% MMS as well as the wild type strain was measured (Figure 22) and each clone was genotyped by PCR (Table 4).

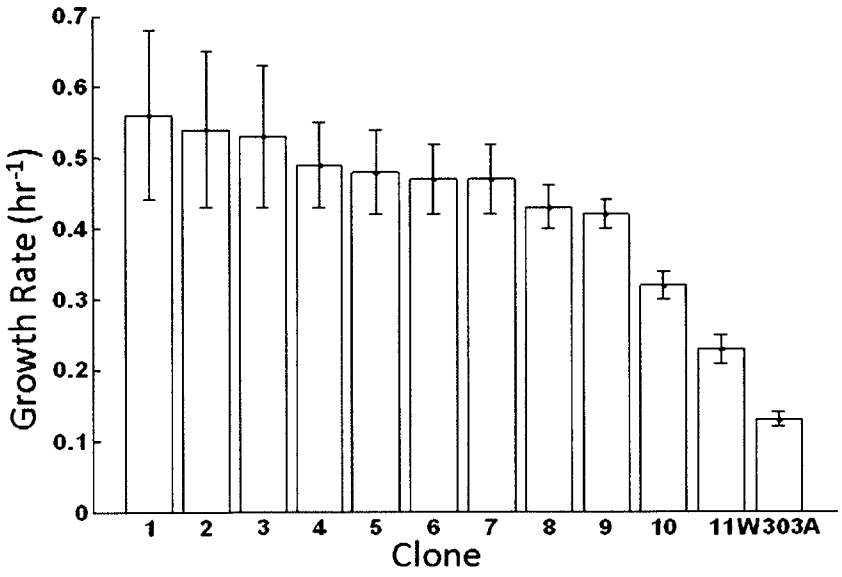


Figure 22: Growth rates of damage-resistant strains in liquid media. Selected clones were able to grow on plates with YPD + 0.035% MMS. Their growth rates in liquid YPD + 0.0035% MMS media were compared to the wild type strain (W303 background). Mean growth rate is plotted. Error bars represent 95% c.i. (based on a student's t-test with three biological replicates).

Table 4: Genotypes and growth rates of selected clones

clone	*APN1*	*APN2*	*REV3*	*REV7*	*ADE2*	A/α	Growth rate[a] (hr^{-1})
1	ON	ON	NAT	NAT	-	A	0.56±0.12
2	NAT	ON	ON	OFF	-	α	0.54±0.11
3	ON	ON	ON	OFF	+	A	0.53±0.1
4	ON	ON	NAT	ON	+	A	0.49±0.06
5	NAT	ON	ON	OFF	-	A	0.48±0.06
6	ON	ON	NAT	ON	-	A	0.47±0.05
7	ON	ON	NAT	ON	+	α	0.47±0.05
8	NAT	ON	ON	OFF	-	α	0.43±0.03
9	ON	ON	NAT	ON	-	A	0.42±0.02
10	ON	OFF	ON	OFF	+	A	0.32±0.02
11	ON	OFF	NAT	ON	-	α	0.23±0.02
W303	NAT	NAT	NAT	NAT	-	A	0.13±0.01
W303	NAT	NAT	NAT	NAT	+	A	0.13±0.02

[a]Growth range is based on the mean and 95% confidence interval from a student's t-test based on three biological replicates.

Nine of the clones exhibited high growth rates in MMS that were indistinguishable from each other and significantly higher than wild type ($p < 0.025$). All these clones had *APN2* expressed constitutively and *APN1* either expressed constitutively or under native regulation. Yeast express *APN2* at low levels during log phase growth, (410 versus 7300 molecules per cell for *APN1* (Ghaemmaghami et al. 2003), and Apn1p is responsible for 97% of AP endonuclease activity in lysate (Boiteux and Guillet 2004). Therefore, these results indicate that AP endonuclease activity is probably rate limiting at the higher AP site loads conferred by MMS treatment and overexpression of *APN2* likely increases the cell's capacity for error-free BER. This is further supported by the lower growth rates of clones 10 & 11 which overexpressed *APN1* but not *APN2*.

We found no clear role for Pol ζ because 4 of the fast growing 9 clones had *REV7* in the OFF state. We never found both *REV3* and *REV7* in the overexpressed ON state, consistent with Pol ζ

overexpression reducing growth as seen with UV (Rajpal, Wu, and Wang 2000). Surprisingly, however, none of the clones were found to be OFF for *REV3* and in every clone which was OFF for *REV7*, *REV3* was ON. Because the function of Rev3p is in its role in forming Pol ζ by dimerizing with Rev7p, it seems unclear whether *REV3* expression is required for fast growth in MMS. Clearly, the *REV3* OFF state is present in the library based on PCR (Figure 20C) and PCR analysis of 20 clones selected from the randomized population plated on rich media found 3 with *REV3* OFF promoter orientations. This phenotype is consistent with a recent finding in mammalian cells which identified a *REV3*-specific role in replication past common fragile sites (Bhat et al. 2013).

The mutagenic conditions and slow growth rates inherent on high levels of MMS had the potential to give rise to resistance phenotypes independent of the randomization of DNA repair genes. However, if unrelated mutations were relevant we would not expect the pattern of genotypes observed in the 11 clones. Furthermore, a clone selected from growth at 0.035% MMS with an *APN1 APN2* overexpression genotype showed no significant difference in tolerance versus a clone with an identical genotype selected under the milder conditions of 0.001% MMS used in Section 3.4 where conditions were used to screen for all (or most) possible genotypes and subsequent mutations could not be responsible for the tolerance. When the two clones were replated on YPD + 0.035% MMS to verify the phenotype, only approximately 25% of plated cells were grew (Figure 23). This plating efficiency is still much greater than wild type cells (Figure 23), but the reason for its reduction is unclear. One possiblility is that the survival phenotype keeps a fraction of the cells alive for long enough for nearby levels of MMS to be depleted or degraded before cells can grow. Alternatively, the phenotype may still be limited in

the tolerance it confers, with a high probability that the damage accumulated in cells will still cause death in most cases.

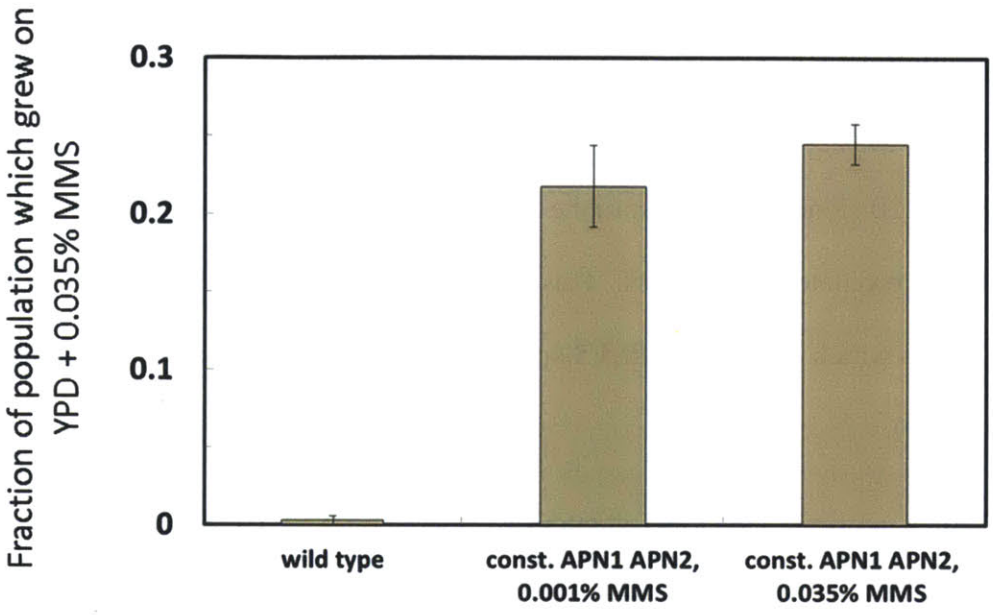


Figure 23: Clones with the constitutively overexpressed *APN1 APN2* expression genotype isolated with different selection pressure have similar tolerances to growth on 0.035% MMS. Comparison of a constitutively active *APN1 APN2* strain (clone 1, Table 4) isolated under 0.035% MMS on YPD to a clone with identical genotype isolated for increased colony size when grown under 0.001% MMS on YPD. Equal numbers of cells (~300) from an overnight liquid YPD culture were plated on YPD and YPD with 0.035% MMS. This plating efficiency represents the fraction of colonies appearing on the plate with MMS, as compared to the plate without. Each clone was done in triplicate, and error bars represent SEM. The tolerance of both clones is not significantly different (2-tail T-test, p = 0.20) but much higher than wild type.

3.4 Discussion

By randomizing the expression of four genes in BER, we rapidly generated a library of 81 genetic variants and found an expression motif which supported growth on high levels of MMS. By examining the genotypes of these variants, we found that overexpression of a BER gene along with the presence of *REV3* (native or constitutively expressed) may increase the capacity for BER repair and support recent findings that identify a role for *REV3* in chromosomal stability which is independent of *REV7* (Bhat et al. 2013).

Understanding how the different modes of repair affect growth fitness and mutation is important in the context of population adaptability. All organisms possess low but non-zero spontaneous mutation rates which are generally neutral or deleterious but could rarely confer a growth advantage. While mutation is necessary to confer population adaptability to changing environments, high rates of mutation can decrease population growth rates, and are sometimes undesirable in stable environments (Clune et al 2008). Mutator strains of bacteria in stress or nutrient poor conditions tend to out-compete wildtype strains (Chao and Cox 1983; Giraud et al. 2001) because they acquire adaptive mutations more quickly, and some bacteria may even possess adaptive systems to raise the spontaneous mutation rate in times of stress (Rosenberg 2001). Knowledge of how particular DNA repair motifs affect mutation and growth in stress conditions may reveal what a mutagenic stress response in eukaryotes would look like and how damage is repaired in these conditions.

In the selection, we found no clones in which both components of Pol ζ were overexpressed. This may be due to either an increased mutagenic burden caused by both MMS and increased error-prone bypass of lesions during replication, or it could be evidence that increasing repair through error-prone pathways is detrimental because it decreases the amount of error-free post-replication repair that occurs, or both. This raises questions about the relationship between growth fitness and mutation rate. A previous study found that overexpressing Pol ζ in UV stress (Rajpal, Wu, and Wang 2000), suggests that as growth rate increases there is a monotonic decrease in mutation rate. However, this does not have to be the case as increasing the flux through error-prone repair could decrease the rate of death (Figure 16), potentially increasing both growth and mutation rates. In the selection experiment in Section 3.3 we did isolate a clone which grew faster in MMS conditions than the wild type. However we originally selected for

cells able to grow in conditions in which all other cells died. By sampling from the entire population in non-selective conditions, we could probe the relationship between growth rate and mutation rate in the context of DNA repair enzymes.

Toward this end, we have already generated a library in which we randomized the expression of the same set of genes (with the addition of *RAD4*). We confirmed diversity in growth rates in the randomized population when grown on rich media with and without low levels of DNA damage (0.001% MMS) and picked 48 clones to represent the range of observed colony sizes (Figure 24 and Figure 25). In future work, we plan to measure the mutation rates of the clones and use their genotypes to find how strains with different modes of repair flux react to mutagenic environments (Table 5).

Comparing growth and mutation rates of individual clones in our generated library under conditions of different abasic site loads would help identify whether both mutation rates and growth rates could increase, and what associated changes in expression of BER genes are required. We might expect all clones to have decreased growth rates and increased mutation rates in the presence of MMS, regardless of the DNA repair environment. This could be a result of a trade-off between mutation rate and growth rate, indicating that mutations arising from error-prone repair are almost always detrimental unless they lead to adaptive mutations. Conversely, we may find that by changing the flux of DNA repair in a way that has not been previously reported that we can increase mutation rate *and* growth rate compared to the wild type background. The latter phenotypic combination would have important ramifications in the design of mutator strains and further understanding the limits of fates of DNA repair.

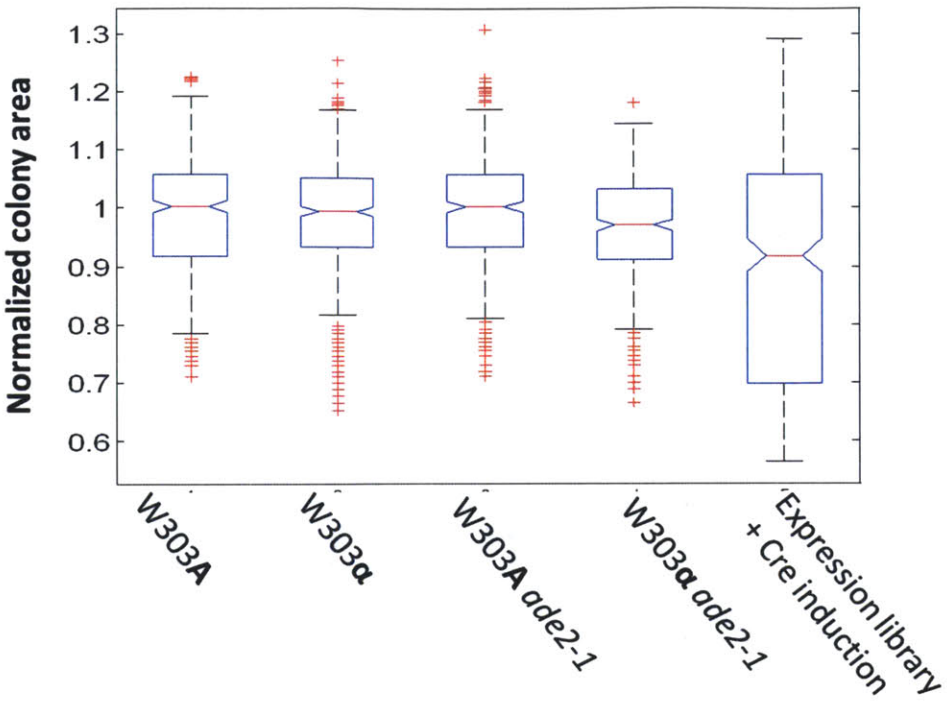


Figure 24: Randomization of five DNA repair enzymes generates growth diversity. Boxplots of colony size distribution after 1 day of growth on plating containing YPAD + 0.001% MMS. Cells were grown in liquid YPAD media for 24 hours before plating. YPAD eliminates differences in growth created but different *ADE2* auxotrophy. N = 421, 294, 284, 228, and 275 colonies, respectively.

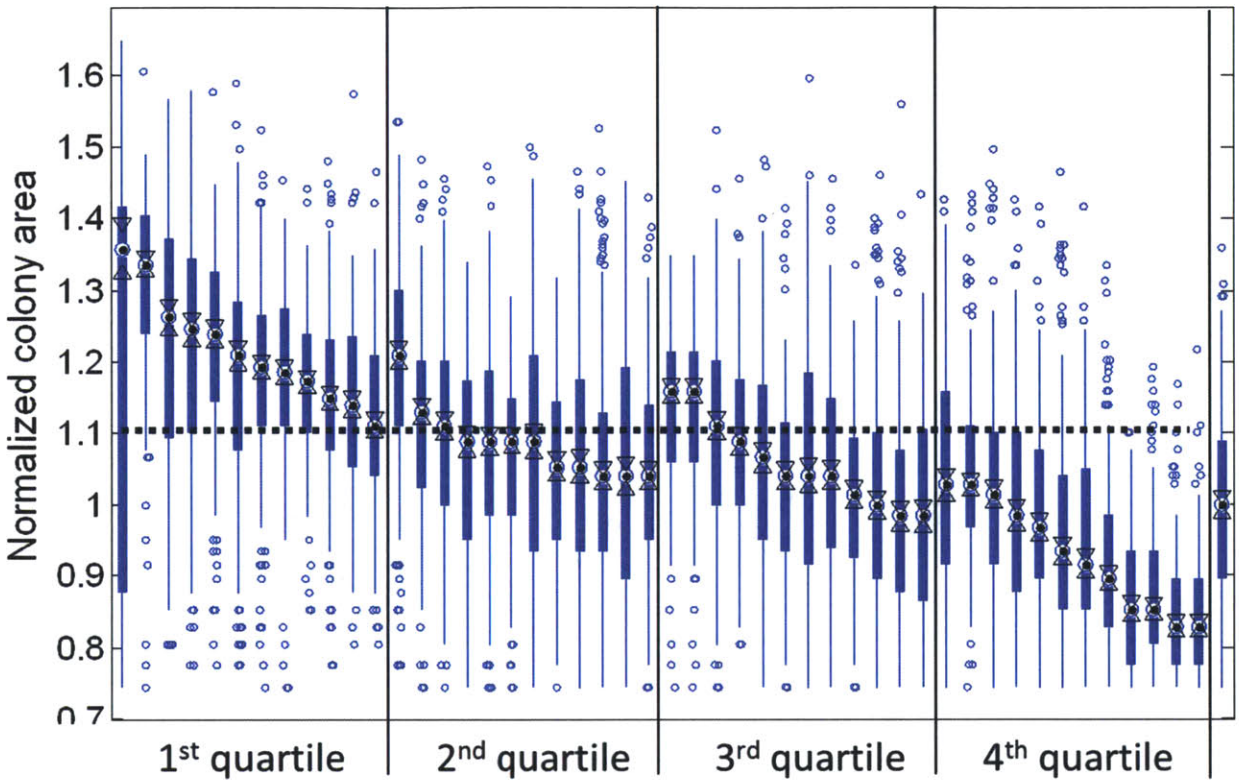


Figure 25: Clonal populations cover a range of growth rates. Boxplots of colony size distributions of each clone after 1 day of growth on YPAD + 0.001% MMS. The clones are grouped according to the quartile in **Figure 24** (far right distribution) from which they were originally selected and plotted in order of descending median. The horizontal dotted line is meant as a guide to the eye and intersects the median of the slowest growing clone picked from the 1st quartile. All plots are normalized to the median of the W303**A** strain grown in YPAD with no MMS (far right). N ~ 400 colonies per data set.

Table 5: Genotypes and median colony sizes of each clone.

clone	*REV3*	*REV7*	*APN1*	*APN2*	*RAD4*	YPAD[a]	MMS[a]	Δ
9	OFF	NAT	ON	NAT	ON	1.2168	1.3579	0.141
4	NAT	OFF	NAT	NAT	NAT	1.3698	1.3352	-0.0346
3	OFF	ON	NAT	NAT	ON	1.1312	1.2623	0.1312
12	NAT	OFF	NAT	NAT	NAT	1.2845	1.2456	-0.0389
2	NAT	OFF	NAT	NAT	OFF	1.1056	1.2386	0.133
10	NAT	OFF	NAT	NAT	ON	1.2092	1.2092	0
24	OFF	NAT	ON	NAT	NAT	1.2168	1.2092	-0.0076
11	NAT	OFF	NAT	NAT	ON	1.2014	1.1934	-0.008
5	NAT	OFF	NAT	NAT	NAT	1.3579	1.1852	-0.1726
47	OFF	NAT	NAT	ON	NAT	1.1109	1.1593	0.0484
38	NAT	OFF	NAT	NAT	NAT	1.054	1.1502	0.0962
43	NAT	NAT	NAT	ON	NAT	1.1212	1.1408	0.0196
15	NAT	ON	ON	NAT	NAT	0.9351	1.1312	0.1961
20	NAT	NAT	NAT	NAT	ON	1.1408	1.1109	-0.0299
33	NAT	OFF	ON	NAT	NAT	0.9351	1.1109	0.1759
16	OFF	ON	ON	NAT	OFF	0.8538	1.1109	0.2572
23	NAT	NAT	ON	NAT	OFF	1.2014	1.0894	-0.112
30	NAT	NAT	ON	ON	NAT	0.985	1.0894	0.1044
6	NAT	NAT	NAT	NAT	ON	1	1.0894	0.0894
45	NAT	NAT	NAT	ON	NAT	1.0663	1.0894	0.0231
22	NAT	NAT	NAT	NAT	NAT	1.1003	1.0894	-0.011
29	ON	NAT	ON	ON	NAT	0.897	1.0663	0.1693
14	NAT	OFF	NAT	NAT	NAT	0.985	1.054	0.0691
8	ON	OFF	NAT	ON	OFF	0.9166	1.054	0.1375
13	ON	NAT	OFF	ON	NAT	1.0414	1.0414	0
17	OFF	NAT	ON	ON	NAT	1	1.0414	0.0414
19	NAT	NAT	NAT	NAT	OFF	0.985	1.0414	0.0564
31	NAT	NAT	ON	NAT	NAT	0.9526	1.0414	0.0888
1	ON	NAT	NAT	NAT	OFF	0.8538	1.0414	0.1876
41	NAT	NAT	OFF	OFF	NAT	1.1109	1.0414	-0.0696
7	NAT	ON	NAT	NAT	NAT	1.1593	1.0281	-0.1312
28	NAT	OFF	NAT	NAT	OFF	1.0894	1.0281	-0.0612
46	NAT	NAT	NAT	OFF	ON	1.2524	1.0144	-0.238
44	NAT	NAT	NAT	NAT	OFF	0.9166	1.0144	0.0978
26	NAT	NAT	NAT	NAT	OFF	0.8865	1	0.1135
18	NAT	NAT	NAT	NAT	NAT	0.9351	0.985	0.0499
34	ON	NAT	OFF	OFF	NAT	1.0894	0.985	-0.1044
42	NAT	NAT	NAT	OFF	OFF	1.0663	0.985	-0.0813
25	NAT	OFF	OFF	OFF	ON	0.8538	0.9692	0.1154
27	NAT	NAT	NAT	NAT	OFF	0.897	0.9526	0.0556
40	NAT	OFF	OFF	ON	NAT	0.9166	0.9351	0.0185
32	ON	NAT	OFF	ON	NAT	1.1408	0.9166	-0.2242
35	NAT	NAT	NAT	NAT	NAT	1.1312	0.8538	-0.2774
37	NAT	NAT	OFF	NAT	OFF	0.8538	0.8538	0
W303A	NAT	NAT	NAT	NAT	NAT	1	0.8538	-0.1462
39	NAT	OFF	OFF	NAT	OFF	0.8039	0.8298	0.0259
48	OFF	NAT	NAT	OFF	NAT	0.8538	0.8298	-0.024

[a]All colony size measurements are the median of the log area, normalized the wildtype control strain on YPAD.

3.5 References

Bhat, Audesh, Parker L. Andersen, Zhoushuai Qin, and Wei Xiao. 2013. "Rev3, the Catalytic Subunit of Polζ, is Required for Maintaining Fragile Site Stability in Human Cells." *Nucleic Acids Research* 41 (4): 2328-2339.

Boiteux, Serge and Marie Guillet. 2004. "Abasic Sites in DNA: Repair and Biological Consequences in Saccharomyces Cerevisiae." *DNA Repair* 3 (1): 1-12.

Chao, Lin and Edward C. Cox. 1983. "Competition between High and Low Mutating Strains of Escherichia Coli." *Evolution*: 125-134.

Clune, Jeff, Dusan Misevic, Charles Ofria, Richard E. Linski, Santiago F. Elena, and Rafael Sanjuan. 2008 "Natural Selection Fails to Optimize Mutation Rates for Long-Term Adaptation on Rugged Fitness Landscapes." *PLOS: Computational Biology* 4(9): e1000187. doi:10.1371/journal.pcbi.1000187

Demple, Bruce and Lynn Harrison. 1994. "Repair of Oxidative Damage to DNA: Enzymology and Biology." *Annual Review of Biochemistry* 63 (1): 915-948.

Ghaemmaghami, Sina, Won-Ki Huh, Kiowa Bower, Russell W. Howson, Archana Belle, Noah Dephoure, Erin K. O'Shea, and Jonathan S. Weissman. 2003. "Global Analysis of Protein Expression in Yeast." *Nature* 425 (6959): 737-741.

Giraud, Antoine, Ivan Matic, Olivier Tenaillon, Antonio Clara, Miroslav Radman, Michel Fons, and François Taddei. 2001. "Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut." *Science* 291 (5513): 2606-2608.

Haracska, Lajos, Ildiko Unk, Robert E. Johnson, Erik Johansson, Peter MJ Burgers, Satya Prakash, and Louise Prakash. 2001. "Roles of Yeast DNA Polymerases Δ and Z and of Rev1 in the Bypass of Abasic Sites." *Genes & Development* 15 (8): 945-954.

Johnson, Robert E., Carlos A. Torres-Ramos, Tadahide Izumi, Sankar Mitra, Satya Prakash, and Louise Prakash. 1998. "Identification of APN2, the Saccharomyces Cerevisiae Homolog of the Major Human AP Endonuclease HAP1, and its Role in the Repair of Abasic Sites." *Genes & Development* 12 (19): 3137-3143.

Memisoglu, Asli and Leona Samson. 2000. "Base Excision Repair in Yeast and Mammals." *Mutation Research* 451 (1-2): 39.

Prakash, Satya, Robert E. Johnson, and Louise Prakash. 2005. "Eukaryotic Translesion Synthesis DNA Polymerases: Specificity of Structure and Function." *Annu.Rev.Biochem.* 74: 317-353.

Rajpal, Deepak K., Xiaohua Wu, and Zhigang Wang. 2000. "Alteration of Ultraviolet-Induced Mutagenesis in Yeast through Molecular Modulation of the< i> REV3 and< i> REV7 Gene Expression." *Mutation Research/DNA Repair* 461 (2): 133-143.

Ramotar, Dindial, Sonya C. Popoff, Edith B. Gralla, and Bruce Demple. 1991. "Cellular Role of Yeast Apn1 Apurinic Endonuclease/3'-Diesterase: Repair of Oxidative and Alkylation DNA Damage and Control of Spontaneous Mutation." *Molecular and Cellular Biology* 11 (9): 4537-4544.

Rosenberg, Susan M. 2001. "Evolving Responsively: Adaptive Mutation." *Nature Reviews Genetics* 2 (7): 504-515.

Xiao, Wei and Barbara L. Chow. 1998. "Synergism between Yeast Nucleotide and Base Excision Repair Pathways in the Protection Against DNA Methylation Damage." *Current Genetics* 33 (2): 92-99.

Chapter 4: Future Directions

We have demonstrated the use of Cre recombinase in randomizing gene expression of both heterologous fluorescent reporters and endogenous yeast genes at disparate loci in an independent manner. With the availability of the plasmids in Figure 4, deployment of the system in yeast requires only a few cloning steps, primer design for disruption of endogenous genes, and sequential integration. Because the Cre recombinase is functional in bacteria and higher organisms (Nagy 2000) extending the system should also be straight forward. Our method can be expanded to target large numbers of genes, making it well-suited to probing complex genetic interactions. Below, we discuss ways to expand the promoter recombination system further, as well as a number of specific gene networks as targets for randomization.

4.1 Overcoming limitations on the number of target genes

For each new floxed promoter the number of genotypes after randomization increases exponentially. However, we have observed factors which will limit the amount of targets. Multiple copies of identical loxP sequences increase the chance of illicit recombination between different floxed promoters. Furthermore, engineering any strain with over a dozen floxed promoters has its own set of challenges associated with using and reusing available markers, which will increase the required number of integration steps. Each subsequent step also becomes more difficult due to increased sequence homology from previous changes.

The use of additional recombinases and recognition sites should decrease irreversible recombination events. Based on probable long range interactions between recombination sites

leading to chromosomal instability, we suggested that no more than 5 two-state promoters or 2 four-state promoters of any loxP variant be used in the same strain, with three orthogonal loxP variants available. Adding new recombinases would decrease the problems caused by having many floxed promoters in the cell. One or two new recombinases, such as engineered versions of Cre or the Flp recombinase from yeast, could allow up to a few dozen two-state promoters in a strain. At least two Cre mutants are known to recombine loxP variants which Cre cannot recombine (Santoro and Schultz 2002). Similar to Cre-loxP, the Flp-FRT system possesses several orthogonal FRT sequence variants (Turan et al. 2010). Multiple recombinases could be expressed from the same plasmid or endogenous locus, with identical inducible promoters. Diverse recombinase/recognition site pairs also prevent intra-chromosomal recombination events caused by randomizing two endogenous genes that are on the same chromosome. Using Cre and Flp, with three recombination site variants per recombinase, we surmise that 30 (5 genes/recombination site) genes could be randomized after 15 integration steps plus mating and then sporulation similar to Chapter 3.

We have outlined a method to do **n** modifications in **n**/2 parallel steps by splitting the integration of floxed promoters across two haploid strains which can then be mated and sporulated. This will alleviate some difficulty when integrating multiple copies of the same floxed promoter, although it is likely that different constitutive promoters will be necessary. For the case of 30 target genes, we suggest using at least 5 constitutive promoters (3 copies per haploid). High efficiency homologous recombination (HR) techniques, such as those that induce double strand breaks (DSBs) at target loci, would make strain building easier. The CRISPR-Cas9 system, discussed in Chapter 1, uses RNAs targeted to specific loci to create DSBs. The guide RNAs (gRNAs) can include multiple target sequences and can induce DSBs at multiple locations in the

genome (Cong et al. 2013). While the CRISPR-Cas9 system has the potential to be used as a multiplex gene targeting tool, even with a DNA oligo library akin to MAGE, the efficiency may be too low to generate the desired amount of diversity. With an efficiency of at most 10%, many integration steps are required to realize a population where a fraction of individuals have experienced multiple successful integrations, especially if using a DNA oligomer library. If multiple rounds of integration are required anyway, they might best be used to sequentially put floxed promoters at specific loci instead of using a multiplexed approach. Afterward, *all* loci are randomized during recombinase induction.

4.2 Branched pathways and signaling networks are ideal systems for randomization

Our randomization system is well-suited to branched networks, where 'flux,' such as chemical intermediates or information signals, is directed to different outcomes. 'Flux' can refer to a range of attributes such as carbon flux though metabolic pathways or the flux of DNA repair when a single damaged base can be repaired through a number of independent pathways. Metabolic engineering already focuses on decreasing carbon flux through undesirable pathways (Alper, Miyaoku, and Stephanopoulos 2005), or overexpressing the enzymes along pathways which lead to valuable products (Wang et al. 2009). Families of potentially high value compounds such as isoprenoids (Armstrong and Hearst 1996; Eisenreich et al. 2004; Fellermeier et al. 2001) are highly branched and specialized, with terpenoids being precursors for other groups of chemicals.

The ability to combine overexpression and deletion (with the two-state promoter) or sample a wide expression range (with the four-state promoter) is also an advantage of our system. Combining deletions of native genes and overexpression of heterologous one can further

improve heterologous pathway production (Koffas, Jung, and Stephanopoulos 2003; Pitera et al. 2007). Tuning expression of genes can be necessary if metabolic pathways impose burdensome metabolic loads on the host (Glick 1995; Parekh and Wittrup 1997), to balance stoichiometry when combining enzymes from different organisms (Dueber et al. 2009).

Finally, signaling pathways monitor multiple stimuli and process that information to mount an appropriate response -- a 'flux' of information. Randomizing the expression of genes involved in these pathways can identify to what extent the overall network behavior is robust to changes in individual nodes, and lead to further studies investigating whether apparent redundancies have more subtle roles in the response. Combining overexpression and deletion can reveal general robustness in signaling pathways, such as the galactose-inducible pathway (Acar et al. 2010), This investigation revealed how the architecture of the network, as well as interaction between the activators and inhibitors, can lead to invariance in response despite changes in gene dosage, offering ideas about the minimal requirements of such systems. Our system could be used to confirm this in other signaling networks. Similar signaling pathways with multiple inputs and specific phenotypic outcomes are those that result in flocculation and pseduohyphal growth (Teunissen and Steensma 1995) or general stress response factors (Gasch et al. 2000).

The ability to rapidly generate a phenotypic library will contribute to the engineering of complex phenotypes. While the suggestions in this section are not exhaustive, the high degree of connectivity in branched and signaling pathways also suggests that the phenotypic landscapes arising from combinatorial expression of component genes will be very complex. Deleting or overexpressing a few genes in tandem may only scratch the surface of how the pathway works, and as Acar, et al. (2009) discovered, a deeper approach, including more than just gene deletion at multiple loci, leads to a clearer picture of how gene networks with robust qualities function or

how to control them. Most of the examples discussed in Chapter 1, as well as here, which involve strains with more than two gene expression modifications do not sample a large number of possible expression combinations. Further, single or double deletion phenotypes may not have predictable effects when combined, as Alper et al. (2005) found when combining deletion mutants which had elevated levels of lycopene production. By investigating many different expression combinations to overproduce a product, they also demonstrated that metabolic pathways can have unanticipated epistatic interactions. Our recombination tool allows the user to generate a library of expression states in nearly the same amount of time it would require for them to build a single strain with a particular expression profile, dramatically expanding the scope of the study.

4.3 References

Acar, Murat, Bernardo F. Pando, Frances H. Arnold, Michael B. Elowitz, and Alexander van Oudenaarden. 2010. "A General Mechanism for Network-Dosage Compensation in Gene Circuits." *Science* 329 (5999): 1656-1660.

Alper, Hal, Curt Fischer, Elke Nevoigt, and Gregory Stephanopoulos. 2005. "Tuning Genetic Control through Promoter Engineering." *Proceedings of the National Academy of Sciences of the United States of America* 102 (36): 12678-12683.

Alper, Hal, Kohei Miyaoku, and Gregory Stephanopoulos. 2005. "Construction of Lycopene-Overproducing E. Coli Strains by Combining Systematic and Combinatorial Gene Knockout Targets." *Nature Biotechnology* 23 (5): 612.

Armstrong, Gregory A. and John E. Hearst. 1996. "Carotenoids 2: Genetics and Molecular Biology of Carotenoid Pigment Biosynthesis." *The FASEB Journal* 10 (2): 228-237.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex Genome Engineering using CRISPR/Cas Systems." *Science* 339 (6121): 819-823.

Dueber, John E., Gabriel C. Wu, G. Reza Malmirchegini, Tae Seok Moon, Christopher J. Petzold, Adeeti V. Ullal, Kristala LJ Prather, and Jay D. Keasling. 2009. "Synthetic Protein

Scaffolds Provide Modular Control Over Metabolic Flux." *Nature Biotechnology* 27 (8): 753-759.

Eisenreich, W., A. Bacher, D. Arigoni, and F. Rohdich. 2004. "Biosynthesis of Isoprenoids Via the Non-Mevalonate Pathway." *Cellular and Molecular Life Sciences CMLS* 61 (12): 1401-1426.

Fellermeier, Monika, Maja Raschke, Silvia Sagner, Juraithip Wungsintaweekul, Christoph A. Schuhr, Stefan Hecht, Klaus Kis, Tanja Radykewicz, Petra Adam, and Felix Rohdich. 2001. "Studies on the Nonmevalonate Pathway of Terpene Biosynthesis." *European Journal of Biochemistry* 268 (23): 6302-6310.

Gasch, Audrey P., Paul T. Spellman, Camilla M. Kao, Orna Carmel-Harel, Michael B. Eisen, Gisela Storz, David Botstein, and Patrick O. Brown. 2000. "Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes." *Science Signaling* 11 (12): 4241.

Glick, Bernard R. 1995. "Metabolic Load and Heterologous Gene Expression." *Biotechnology Advances* 13 (2): 247-261.

Koffas, Mattheos AG, Gyoo Yeol Jung, and Gregory Stephanopoulos. 2003. "Engineering Metabolism and Product Formation in< i> Corynebacterium Glutamicum by Coordinated Gene Overexpression." *Metabolic Engineering* 5 (1): 32-41.

Nagy, Andras. 2000. "Cre Recombinase: The Universal Reagent for Genome Tailoring." *Genesis* 26 (2): 99-109.

Parekh, Rajesh N. and K. Dane Wittrup. 1997. "Expression Level Tuning for Optimal Heterologous Protein Secretion in Saccharomyces Cerevisiae." *Biotechnology Progress* 13 (2): 117-122.

Pitera, Douglas J., Chris J. Paddon, Jack D. Newman, and Jay D. Keasling. 2007. "Balancing a Heterologous Mevalonate Pathway for Improved Isoprenoid Production in< i> Escherichia Coli." *Metabolic Engineering* 9 (2): 193-207.

Santoro, Stephen W. and Peter G. Schultz. 2002. "Directed Evolution of the Site Specificity of Cre Recombinase." *Proceedings of the National Academy of Sciences* 99 (7): 4185-4190.

Teunissen, AWRH and HY Steensma. 1995. "The Dominant Flocculation Genes of Saccharomyces Cerevisiae Constitute a New Subtelomeric Gene Family." *Yeast* 11 (11): 1001-1013.

Wang, Harris H., Farren J. Isaacs, Peter A. Carr, Zachary Z. Sun, George Xu, Craig R. Forest, and George M. Church. 2009. "Programming Cells by Multiplex Genome Engineering and Accelerated Evolution." *Nature* 460 (7257): 894-898.

Appendix

Materials and methods

Strain and plasmid construction

Yeast strains were constructed using standard techniques (Guthrie and Fink 2004). All yeast strains used here are in the W303 background (Thomas and Rothstein 1989) and detailed in Supplemental Table I. Plasmids and construction notes are listed in Supplemental Table II. The yeast integration vectors used were linearized by digestion with *Hin*dIII, *Pst*I, and *Eco*RV in the *HIS3, URA3,* and *TRP1* marker regions respectively (Sikorski and Hieter 1989). The PCR-based promoter replacement cassettes based on those designed by (Longtine, et al. 1998) were amplified using the sequences 5' – TGACGTGCGCAGCTCAGGGG – 3', and 5' – ATCGATGAATTCGAGCTC – 3' on the 3' end of sense and antisense primers, respectively. Primers contained approximately 50 bp of homology to the 5' ends to the desired genomic locus.

Flow cytometry

Flow cytometry was performed using a LSRII HTS (Becton-Dickinson). Venus, tdTomato, and Cerulean proteins were detected using a 488-nm laser and 530/30-nm filter, a 561-nm laser and 610/20-nm filter, and a 405-nm laser and 450/50-nm filter, respectively. For each sample, ~100,000 cells were measured and visualized on a contour plot of forward versus side scatter. Cells within a contour containing 10% of the population were used for further analysis.

'Floxed' promoter randomization

Cells containing 'floxed' promoter and galactose-inducible Cre recombinase were grown in 2% raffinose liquid media overnight, up to an optical density at 600 nm (OD_{600}) of ~1. They were then inoculated into 2% galactose media at an OD_{600} of 0.1-0.2 and allowed to grow for 6 hours for randomization unless otherwise noted in the text. Finally, cells were inoculated in glucose media overnight (>18 hrs) to inhibit Cre expression and allow fluorophores to reach steady-state expression levels. The galactose-inducible Cre recombinase was expressed from a centromeric *LEU2*-marked plasmid (kind gift from Paul Wiggins). The centromeric plasmid was maintained in the population by selecting for leucine autotrophy at each step.

Sporulation and haploid selection

Diploid cells heterozygous at the *CAN1* locus (*CAN1/can1*) were sporulated by overnight growth in rich liquid yeast peptone (YP) media with 1% acetate followed by washing and suspension in 1% potassium acetate + complete amino acids for two days. Haploids were selected by plating the sporulated population on synthetic defined media with 600 mg/L canavanine, and lacking arginine and leucine.

Gowth rate measurements

Growth rates in liquid media were measured using 96 deep well plates agitated at 30℃. Three biological replicates per clone were inoculated at 0.003 OD_{600} in 800 μL rich media (yeast peptone with 2% glucose or YPD) or YPD + 0.0035% MMS and grown to early exponential phase (~0.1 OD_{600}), at which point optical density was measured approximately once per hour. OD_{600} data points between 0.1 and 1 were fit well to simple exponential growth, which was used to determine the growth rate. Optical density measurements were taken with a Varioskan Flash (ThermoScientific).

Scoring colony size

We measured colony size in the library as a measure of diversity in growth rate. Briefly, approximately 400 cells were plated on 15 cm plates, allowed to grow 24 hours post-plating, and then imaged using a digital camera. Metamorph (Molecular Devices) software was used to segment and record the size of colonies on the plate.

References

Guthrie C, Fink GR. 2004. Guide to yeast genetics and molecular and cell biology. Part A. Amsterdam:Elsevier Academic Press.

Longtine MS, McKenzie A,3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast 14:953-961.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics 122:19-27.

Thomas BJ, Rothstein R. 1989. Elevated recombination rates in transcriptionally active DNA. Cell 56:619-630.

Strain list

Strain	Relevant Genotype	Parent Strain
Y1	MATa *trp1-1 can 1-100 leu2-3, 112 his3-11, 5 ura3 GAL+*	W303
Y2	MATα *trp1-1 can 1-100 leu2-3, 112 his3-11, 5 ura3 GAL+*	W303
YBN1	MATa *his3*::loxP-P_{TEF1}-loxP-venus-*HIS3*	Y1
YBN2	MATa *ura3*::loxP- P_{TEF1}- loxP-cerulean-*URA3*	Y1
YBN3	MATa *his3*::loxP-P_{TEF1}-loxP-tdTomato-*HIS3*	YBN2
YBN4	MATa *trp1*::loxN-P_{TEF1}-loxN-venus-*TRP1*	YBN3
YBN5	MATα *ura3*::loxP- P_{TEF1}- loxP-cerulean-*URA3*	Y2
YBN6	MATα *his3*::loxP-P_{TEF1}-loxP-tdTomato-*HIS3*	YBN5
YBN7	MATα *trp1*::loxN-P_{TEF1}-loxN-venus-*TRP1*	YBN6
YBN8	MATa/MATα	YBN7, YBN4
YBN9	MATa *ura3*::loxP- P_{ADH1}-$P_{CYC1TATA}$- loxP-loxP-P_{MYO2}-P_{ACT1}-loxP-venus-*URA3*	Y1
YBN10	MATα *ura3*::loxP- P_{ADH1}-$P_{CYC1TATA}$- loxP-loxP-P_{MYO2}-P_{ACT1}-loxP-tdTomato-*URA3*	Y2
YBN11	MATa/MATα	YBN9, YBN10
YBN12	MATa/MATα	YBN6, YBN4
YBN13	MATa *trp1*::loxP-P_{TEF1}-loxP-venus-*TRP1*	YBN3
YBN14	MATa *ura3*::loxP-P_{TEF1}-loxP-venus-*URA3*	Y1
Y343	MATa *ade2-1, trp1-1, leu2-3, 112 his3-11, 5 ura3, GAL+*	W303
YBN15	MATa *APN1*::*hisMX6*-loxP-P_{ACT1}-loxP	Y343
YBN16	MATa *REV7*::*kanMX6*-loxP-P_{TEF1}-loxP	YBN15
YBN17	MATα *APN2*::*hisMX6*-loxP-P_{ACT1}-loxP	Y2
YBN18	MATα *REV3*::hisG-*klURA3*-hisG-loxP-P_{TEF1}-loxP	YBN17
YBN19	MATa/MATα	YBN16, YBN18

Plasmid list with construction details

Plasmid	Base Vector	Attributes	Construction Information
p227	pBS49	(Sauer 1987)	A gift from Paul Wiggins
pBN1	pRS304	loxP-P_{TEF1}-loxP	Amplification of native *S. cerevisiae TEF1* promoter by primers TEFpr loxP + and TEFpr loxP - XhoI/EcoRI digestion and ligation into pRS304
pBN2	pRS304	YFP-venus	EcoRI/SacI digestion of YFP ORF from laboratory stock and ligation into BN1
pBN3	pRS303	YFP-venus	EcoRI/SacI digestion of YFP ORF from BN2 and ligation into pRS303
pBN4	pRS303	loxN-P_{TEF1}-loxN	Amplification of native *S. cerevisiae TEF1* promoter by primers TEF1loxN+ and TEF1loxN- XhoI/EcoRI digestion and ligation into pRS303
pBN5	pRS304	RFP-tdTomato	EcoRI/SacI digestion of RFP ORF from laboratory stock and ligation into BN2
pBN6	pRS306	loxP-P_{TEF1}-loxP	XhoI/EcoRI digestion of 'floxed' *TEF1* promoter from BN4 and ligation into pRS306
pBN7	pRS306	CFP-cerulean	EcoRI/SacI digestion of CFP ORF from laboratory stock and ligation into BN6
pBN8	pRS306	loxP-P_{ADH1}-P_{min}-loxP	Amplication of native *S. cerevisiae ADH1* and *CYC1* promoters with primers PEL ADH1+/PEL ADH1- and PEL CYC1+/PEL CYC1- respectively, fusion of DNA fragments then XhoI/EcoRI digestion and ligation into pRS306
pBN9	pRS306	YFP-venus	EcoRI/SacI digestion of YFP ORF from BN2 and ligation into BN8
pBN10	pRS306	RFP-tdTomato	EcoRI/SacI digestion of RFP ORF from BN5 and ligation into BN9
pBN11	pRS306	loxP-P_{MYO2}-P_{ACT1}-loxP	Amplication of native *S. cerevisiae MYO2* and *ACT1* promoters with primers MYO2 PEL+/MYO2 PEL- and ACT1 PEL+/ACT1 PEL- respectively, fusion of DNA fragments then EcoRI digestion and ligation into BN9
pBN12	pRS306	loxP-P_{MYO2}-P_{ACT1}-loxP	EcoRI digestion from BN11 and ligation into BN10
pBN13	pRS306	YFP-venus	EcoRI/SacI digestion of YFP ORF from BN2 and ligation into BN11
pBN14	pRS306	RFP-tdTomato	EcoRI/SacI digestion of RFP ORF from BN5 and ligation into BN13
pBN15	pRS306	loxP-P_{ACT1}-loxP	Amplification of native *S. cerevisiae ACT1* promoter by primers ACT1loxP1 + /ACT1loxP1 - XhoI/EcoRI digestion and ligation into BN13
pBN16	pRS306	loxP-P_{MYO2}-loxP	Amplification of native *S. cerevisiae MYO2* promoter by primers MYO2loxP1 + /MYO2loxP1 -

			XhoI/EcoRI digestion and ligation into BN13
pBN17	pRS306	loxP-P_{ADH1}-loxP	Amplification of native *S. cerevisiae ADH1* promoter by primers ADH1loxP1 + /ADH1loxP1 - XhoI/EcoRI digestion and ligation into BN13
pBN18	pRS306	loxN-P_{TEF1}-loxN	Amplification of native *S. cerevisiae TEF1* promoter by primers TEF1loxN+/TEF1loxN - XhoI/EcoRI digestion and ligation into BN15
pBN19	pRS306	lox2272-P_{TEF1}-lox2272	Amplification of native *S. cerevisiae TEF1* promoter by primers TEF1lox2+/TEF1lox2 - XhoI/EcoRI digestion and ligation into BN15
pBN20	pRS303	loxN-P_{TEF1}-loxN	XhoI/SacI digestion of 'floxed' *TEF1* promoter from BN18 and ligation into BN3 (loxN variant)
pBN20	pRS303	lox2272-P_{TEF1}-lox2272	XhoI/SacI digestion of 'floxed' *TEF1* promoter from BN19 and ligation into BN3 (lox2272 variant)
pBN21	pRS304	loxP-P_{MYO2}-P_{ACT1}-loxP/ loxP-P_{ADH1}-P_{min}-loxP	XhoI/SacI digestion of promoter and ORF cassette from BN12 into pRS304
pBN22	pFα6-kanMX6 (Wach et al, 1997)	loxP-P_{TEF1}-loxP	Amplication of 'floxed' *TEF1 promoter* from BN1 using primers loxP SacI +/loxP SacI-, SacI digestion and ligation into pFα6-kanMX6
pBN23	pFα6-hisMX6 (Wach et al, 1997)	loxP-P_{TEF1}-loxP	SacI digestion of 'floxed' *TEF1* promoter from BN22 and ligation into pFα6-hisMX6
pBN24	pFα6-natMX6 (Hentges et al, 2005)	loxP-P_{TEF1}-loxP	SacI digestion of 'floxed' *TEF1* promoter from BN22 and ligation into pFα6-hisMX6
pBN25	pFα6-kanMX6	loxP-P_{ACT1}-loxP	Amplication of 'floxed' *ACT1* promoter using primers loxP SacI +/loxP SacI-, SacI digestion and ligation into pFα6-kanMX6
pBN26	pFα6-hisMX6	loxP-P_{ACT1}-loxP	SacI digestion of 'floxed' *ACT1* promoter from BN25 and ligation into pFα6-hisMX6
pBN27	pFα6-natMX6	loxP-P_{ACT1}-loxP	SacI digestion of 'floxed' *ACT1* promoter from BN25 and ligation into pFα6-hisMX6
pBN28	pFα6-kanMX6	loxN-P_{TEF1}-loxN	Amplication of 'floxed' *TEF1 promoter* from BN1 using primers loxN SacI +/loxN SacI-, SacI digestion and ligation into pFα6-kanMX6
pBN29	pFα6-kanMX6	lox2272-P_{TEF1}-lox2272	Amplication of 'floxed' *TEF1 promoter* from BN1 using primers lox2 SacI +/lox2 SacI-, SacI digestion and ligation into pFα6-kanMX6
pBN30	pFα6-hisMX6	loxN-P_{TEF1}-loxN	SacI digestion of 'floxed' *TEF1* promoter from BN28 and ligation into pFα6-hisMX6 (loxN variant)
BN31	pFα6-hisMX6	lox2272-P_{TEF1}-lox2272	SacI digestion of 'floxed' *TEF1* promoter from BN29 and ligation into pFα6-hisMX6 (lox2272 variant)