Low escape-rate genome safeguards with minimal molecular perturbation of Saccharomyces cerevisiae

Low escape rate genome safeguards with minimal molecular perturbation of *Saccharomyces cerevisiae*

Authors: Neta Agmon[1,2], Zuojian Tang[1], Kun Yang[2], Ben Sutter[3], Shigehito Ikushima[2,4], Yizhi Cai[2,5], Katrina Caravelli[2], James A. Martin[1,2], Xiaoji Sun[1], Woo Jin Choi[2], Allen Zhang[2], Giovanni Stracquadanio[2], Haiping Hao[2], Benjamin Tu[3], David Fenyo[1], Joel S. Bader[2], Jef D. Boeke[1,2*]

1. Institute for Systems Genetics, NYU Langone University School of Medicine, 430 East 29th Street, ACLSW Room 562, New York 10016, USA
2. School of Medicine, Edward D. Miller Research Building, Johns Hopkins University, 733 N. Broadway, Baltimore, MD 21205, USA
3. Department of Biochemistry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9038, USA
4. Central Laboratories for Key Technologies, KIRIN Company Limited, Yokohama, Kanagawa 236-0004, Japan
5. School of Biological Sciences, The King's Buildings, University of Edinburgh, Scotland, EH9 3JR, UK

*Corresponding author. E-mail: jef.boeke@nyumc.org (J.D.B.)

Keywords: genome safety; escape mutants; RPD3L; histone deacetylase; yeast

With the increasing use of genetically modified organisms there is a growing need for biocontainment in order to secure biosystems from outside malice as well as to prevent propagation in an open system in the event of advertent or inadvertent release to the environment. Although recombinant DNA technology was established more than four decades ago (1) the fact that genetically modified organisms have not caused any substantial deleterious incident is due in part to precautionary measures taken by professional genetic engineers and the high cost of genetic engineering, largely restricting its use to academic and industrial labs. However, in the current era of DIY synthetic biology, with rapid technological advances in DNA synthesis and assembly (2), genome editing (3, 4) and computational tools for design (5-7), and with the decreasing cost of DNA synthesis, the need for genomic safeguards is clear.

Early biocontainment efforts focused on the use of metabolic auxotrophy dependence (8, 9), toxin/antitoxin dependent suicide (10-13) or both (14, 15). Although top performers using these strategies do comply with the NIH standard for safeguards (10^{-8} escape rate), they are at risk for cross-feeding and the need for micromolar concentrations of nutrients, making them costly for industrial scale-up. Recent work has focused on recoded *Escherichia coli* with synthetic auxotrophy to a non-natural amino acid (16, 17), showing robust growth maintained for generations as well as an escape rate of less than 10^{-11}. In addition, previous work from our lab has shown engineering of a yeast strain carrying multiplex biocontainment combining transcriptional and recombinational based safeguards, reaching an escape rate of less than 10^{-10} (18). This was achieved in part with transcription control of a single essential gene with an escape rate of 10^{-6}, however this could potentially be lowered by fine-tuning of the safeguard construct, reducing the multiplex escape rate to a much lower value. Here we focus on develop such enhanced transcriptional safeguards that are about an order of magnitude more “secure” individually. Also, analysis of the escapers provides insights into their molecular basis and potentially can help engineer reduced escape rates in the future. Finally, there may be many other generic uses for tightly controlled essential genes in fundamental genetic studies.

The yeast *Saccharomyces cerevisiae* is a critically important host for fermentation and production of many biomedically and industrially critical products (19). In our attempts to find best-acting transcriptional safeguard(s) for yeast cells we sought a construct achieving the following criteria: fitness approaching the wild type in permissive conditions, lowest possible escape rate under restrictive conditions and essential growth supplement in a nanomolar scale. We screened a library of essential genes to find the best acting genes. Using combinatorial assembly we screened for the best acting promoters and terminators, and finally, we used further genetic engineering to fine-tune safeguard constructs.

In addition, we examined components that might be employed in a setting where the aim is to prevent the theft of intellectual property in the form of a living strain. To this end, we performed an analysis of harmless but ineffective regulatory genes and “decoy molecules” which could be employed to mask potential proprietary components used in

safeguarded strains and the medium on which they depend in an industrial setting, to discourage their reverse engineering.

Results

Broad screen for candidate essential genes to be incorporated into safeguards

Our previously described transcriptionally regulated safeguards (SGs) consist of a promoter regulated by an externally supplied ligand, driving the expression of one or more essential genes (18). For these to meet the requirements of 1) robust growth in the presence of the ligand and 2) complete failure to grow or loss of viability in the absence of ligand, we sought yeast genes best adapted to the purpose. Such genes are expected to tolerate a certain variation in their expression without a fitness penalty, but stop growing entirely, or lose viability when expression falls below a certain threshold- i.e. they should not display "leaky growth" when minimally expressed. To identify potential safeguard essential genes we began by identifying the 250 strains from the Yeast Tet-Promoters Hughes Collection (yTHC) annotated as having a "Severe" or "Very Severe" phenotype on Doxycycline in a "Tet-OFF" strain background (20). These 250 strains, as well as a WT control (Strain R1158) were grown in YPD medium, diluted serially and plated on YPD, YPD+G418 (selecting for the integrated Tet promoter-essential gene expression cassette) and YPD+10 µg/mL Doxycycline. All strains grew as expected both on YPD and on YPD+G418, but, as expected, showed various colony sizes (fitness) in the presence of doxycycline. In addition, the different strains also showed various fitness levels in the absence of doxycycline. From the 250 strains initially chosen for this analysis we chose to continue with a subset of 47 strains, all of which showed high fitness (similar to WT) on YPD, and low fitness in the presence of Doxycycline (Supplementary Figure S1).

Building shuffle strains for combinatorial screening of safeguard strains

Because our safeguard strategy relies on using essential genes, we must introduce the safeguard transcription units (SGTUs) into a yeast strain that already expresses the gene of interest. Having the "unguarded" copy of the gene on a *URA3* "shuffle plasmid" allows quick and efficient spontaneous plasmid loss in the presence of ligand, leaving only the safeguarded copy. Thus we constructed 49 haploid "shuffle strains", carrying a deletion of each essential gene as well as a shuffle plasmid with a wild type copy of the essential gene and a *URA3* marker, enabling its loss using 5-Foa counter selection (21).

To create shuffle plasmids for each candidate gene we amplified the corresponding CDS (coding sequence) from the genome with 500bp upstream and 200bp downstream sequences, to include the native promoter and terminator sequences. The primers used for the genome amplification include 50bp overhangs, which served as recombination sites with pRS416 (Figure 1). Each PCR product was co-transformed with linearized pRS416 [a centromeric *URA3* plasmid (22)] into a diploid strain heterozygous for the corresponding essential gene. Homologous recombination between the PCR product and the vector facilitated construction of a shuffle plasmid containing *URA3* and the essential gene *in vivo* (Figure 1). These Ura^+ diploids were sporulated and dissected to isolate "shuffle strains": haploids containing the shuffle plasmid and lacking the essential gene

of interest. Of the 49 candidates we successfully constructed shuffle strains for 45 potential SGTUs.

Combinatorial yGG assembly of safeguard constructs

Having the ability to combinatorially assemble SGTUs provides many more variants for testing than could be made individually. Thus, we assembled our SG constructs as a combinatorial yeast Golden Gate (yGG) assembly (23) and screened for the best candidates in yeast. We amplified all the parts (promoters, essential gene CDSs, and terminators) from the yeast genome with the appropriate yGG overhangs, as described (23). For each essential gene we performed a "one pot" yGG assembly, adding 6 distinct galactose-regulated promoters [*GAL1, GAL7, GAL10, SPAL2, SPAL5 and SPAL6* (24)], *GAL1* terminator and acceptor vector (pAV10.HO5.loxP). The *SPAL* promoters were of special interest because earlier work showed that these promoters greatly reduced the leakiness of the associated CDSs (24); the number following the name SPAL indicates the number of GAL4 recognition sequences in the promoter. Following yGG assembly, the reaction mix was transformed into bacteria and grown in liquid medium, producing a bulk plasmid prep which was digested to evaluate assembly efficiency. Assembly efficiency is evaluated by digesting the bulk plasmid prep with *Not*I (flanking the assembled SGTU and the homologous arms for integration) which revealed a band corresponding to the size of the assembled TU (data not shown).

Screening for candidate safeguard strains

Following combinatorial yGG assembly and verification, each assembly was transformed into its corresponding yeast shuffle strain. Transformations were plated on SGal–Leu to allow expression of the promoters and enable loss of the shuffle plasmid. Transformed colonies were then replica plated to SC+5-Foa 2% galactose plates. In most cases the majority of the colonies were 5-Foa resistant. 10 Foa^R isolates were chosen from each gene and were plated on YPD and YPGAL to assess ability to serve as a safeguard strain (Supplementary Figure S2). Note that there are significant differences between genes as well as between variant TUs containing the same CDS, presumably reflecting distinct promoter characteristics. The 12 candidate genes were chosen that perform best (they grew well on YPGAL and showed low background growth on YPD; Figure 2A). 8 isolates from each of 12 essential genes were chosen, and subsequently transformed with pRS416-GEV, which expresses the GAL4-estrogen binding domain-VP16 transcriptional activator, and is activated by low concentrations of estradiol (25). Each strain was transformed with empty vector as a negative control. The optimal SG candidates should grow well with Estradiol (SC–Ura + estradiol) and should show a very low or no growth on SC–Ura without estradiol. Figure 2B shows the best 5 strains out of the 12 essential genes plated on SC–Ura+ 30nM Estradiol vs. SC–Ura. As shown in Supplementary Figure S3, we had a large number of "good" candidate strains to choose from. We continued after selecting 3 "good" strains from each of the 5 SGs (Except for *RPB11,* which had only one good strain).

All 13 strains were examined for doubling time in SC–Ura + 30nM estradiol. They were all compared to their corresponding shuffle strain (Figure 2C). Similarly to the previously described histone H3/H4 SG strain (18), all SG strains show a slight increase in doubling

time compared to the shuffle strain, which we attribute to the influence of induction of endogenous *GAL* genes by estradiol (26). Of the 13 we chose one SG strain for each essential gene to follow up with an escape (reversion) rate experiment, as a final test for optimality, and identification of the promoters.

In addition, comparison of the three best SG strains identified here with the previously published ones (18) shows significantly lower background growth on YPD medium compared to YPGAL (Figure 2D), indicating that this screen enabled selection of better performing SG strains.

Measuring fitness of safeguard strains

Due to the fact that these strains were created in a combinatorial assembly strategy and chosen from a pool of safeguards, we decided to evaluate performance using individually constructed plasmids, and to evaluate robustness in distinct genotypic backgrounds, we tested in both *MAT**a*** and *MAT**α*** host strains. This was done using yeast Golden-Gate (yGG) assembly transformed into yeast as described above. Candidates were plated on medium with or without 30nM Estradiol for verification. The results were similar to the counterparts made using combinatorial assembly (data not shown) in both mating types. To evaluate fitness of the strains we measured their growth rate and calculated their doubling time compared to WT strain carrying the GEV plasmid (Figure 3A). Growth rate in all three safeguard strains shows little to no difference from the control strain.

In order to examine the difference between safeguard strains and control we preformed transcriptome and metabolomics profiling (Figure 3, Supplementary Table S1). Transcriptome analysis showed that in most of the SG strains there are only 0-7 genes which showed a significant change in transcript level (Table S1). Only one SG strain, *RPB11 MAT**a***, showed a larger number of transcripts that changed significantly compared to the control strain.

For metabolomics profiling we compared each strain to a WT strain carrying the GEV plasmid. Figure 3C shows a heat-map of the different metabolites and their fold difference compared to their control (Supplementary Table S2). In all strains (both *MAT**a*** and *MAT**α***) there was almost no detectable difference compared to the control stain. In agreement with the transcriptome and growth rate analyses these safeguard strains show little to no differences from the control strain.

Importantly, fitness was measured compared to a control carrying the GEV expression plasmid, this was done based on a previous observation that expression of the GEV construct causes a global activation of Gal-UAS regulated genes and thus reduces growth rate of the strain (26). A solution to this limitation is discussed below.

In addition, we measured the escape rate of our SG strain, all 5 SG strains had less than 10^{-7} escapers per cell division (Table 1).

Analysis of escapers (revertants)

Escape rate experiments were performed on our safeguard strains; 16 independent escapers were isolated from each strain (8 *MAT**a*** and 8 *MAT*α). For complementation group analysis, strains were replica-plated in a crisscross pattern to allow for mating at the junctions. Mating was performed on SC–Ura permissive medium containing estradiol, and then replica plated to SC–Met–Lys–Ura + estradiol selecting for diploids and finally replica plated to SC–Met–Lys–Ura without estradiol for analysis of complementation groups. Escapers that originated via a recessive mutation in the same gene fail to complement and thus grow on medium lacking estradiol. In contrast, escapers originating from recessive mutants in different genes will complement and fail to grow on medium lacking estradiol. In addition, backcrosses to the original SG strain indicate whether the mutation is recessive or dominant. Table 2 summarizes the complementation analyses. We isolated 8 complementation groups for *FAS2* SG escapers and 7 complementation groups for *SEC4* SG escapers. For *RPB11*, all escapers belong to the same complementation group, excepting R3 which showed a dominant phenotype. One candidate from each complementation group was selected for further analysis.

In order to analyze the location (GEV plasmid or genome) of the mutation in each escaper, we isolated the plasmid from each escaper strain analyzed and sequenced it. Digestion of plasmid DNA revealed that all but one (*RPB11 MAT**a*** R1) showed a similar plasmid digest pattern as the original plasmid (Supplementary Figure S4). Sequencing of the plasmid isolated from *RPB11 MAT**a*** R1 showed a 684bp deletion in the Estrogen Binding domain (EBD). This explains both the growth in the absence of estradiol due to constitutive localization in the nucleus and well as the escaper's dominant phenotype.

To locate the mutations in the remaining escapers we performed the following: 1) Cured each escaper of its GEV plasmid; 2) Transformed the isolated plasmids into the original corresponding safeguard strain and 3) Transformed a WT plasmid (pRS416-GEV) into the cured escaper (point 1).

All of these were plated as serial dilutions on SC–Ura supplemented with 2% Galactose, SC–Ura + 30nM Estradiol and SC–Ura (Supplementary Figure S5). In all but one strain (*RPB11 MAT**a*** R1), these experiments indicated that the mutation was in the genome, as transforming the isolated plasmid from the escapers did not confer the capability to grow without estradiol, however, the cured strain transformed with the original WT plasmid (pGEVnew) reconstituted the escaper phenotype.

We also analyzed the escapers cured of the plasmid on YPGAL medium (Supplementary Figure S6). Interestingly, all escapers showed a significantly slower growth rate in YPGAL compared to the original SG, conferring a fitness disadvantage compared to the original SG strains. This was also observed (to a lesser extant) when escapers were plated on SC–Ura + 30nM Estradiol (Supplementary Figure S5), which conferred a fitness disadvantage relative to the original SG strains .

In order to locate the genomic mutations in these escapers we performed whole genome sequencing (one from each complementation group). In all but one, *RPB11* R3, genome

sequencing revealed mutations in one of the five components of the Rpd3L complex (Table 3). This complex represses transcription of URS1-containing genes by deacetylation of histone H3 and H4 in the promoter (27). Due to the fact that all of our high-performing SG constructs are driven by a SPAL promoter (24) which contains the URS1 sequence to minimize background expression (28), mutations in genes encoding the components of the Rpd3L complex are predicted to alleviate the repression and cause expression of the SG gene even in the absence of estradiol. In *RPB11* R3, there was a single bp mutation of the URS1 sequence (AGCCGCCGA to AGCTGCCGA) which presumably prevents binding of Rpd3L to the promoter and explains the dominant phenotype of this escaper. The identification of the mutations also explains the fitness defect in the permissive media; these mutants are known to have intrinsic slow growth defects. Since 100% of the mutations we isolated in *SEC4* and *FAS2* SG strains were recessive, we expect that the escape rate could be lowered substantially by use of diploid host strains. Thus, by crossing the SG strains from the two mating types we engineered diploid SG strains for both *FAS2* and *SEC4*. In both cases the diploids shows reduced escaper rate in media without Estradiol (Figure 4A).

To verify the effect that mutations in Rpd3L component genes have on SG strains, we deleted *UME6* in the *FAS2 MAT**a*** SG strain. As a control, we deleted a member of the Rpd3S complex, *EAF3*. As seen in Figure 4B, deletion of *UME6* but not *EAF3* completely recapitulated the escaper phenotype, including a slight growth defect in medium with estradiol.

Further improvements to safeguard strains.

As mentioned above, expression of the GEV construct causes a modest fitness defect thought to be due to off-target effects due to expression of UAS containing promoters in the genome. This was previously reported by McIsaac and colleagues (26) and a possible alternative to the GEV system was suggested by the same group, namely the ZEV system (26). In the ZEV system the transcription factor (VP16) is fused to an Estrogen-Binding-Domain (EBD) and a Zinc-finger binding domain recognizing a 9 bp sequence that was cloned into a minimal *GAL1* promoter. The ZEV system was shown to have no effect on growth rate in yeast cells presumably due to the absence of native binding sites for this zinc finger protein. Based on the much lower background of the *SPAL* promoter compared to native *GAL* promoters, we decided to clone the Z4 array containing 6 binding sites into the *SPAL* promoter , replacing the *GAL4* recognition sites (Figure 5A and Supplementary Data). This promoter is expected to be as tight as the *SPAL* promoter due to its URS1 sequence (transcribing only in the presence of estradiol and the ZEV protein) and as specific as the ZEV system (no off-target effect). We called the new chimeric promoter the “SPAZ” promoter.

Following the synthesis of the SPAZ promoter we used it to drive expression of the 3 best-performing SG genes (*FAS2, RPB11* and *SEC4*). Serial dilutions on media with or without 1 μM estradiol showed that in all three cases growth with estradiol is indistinguishable from that of the shuffle strain. There was a significant reduction in viability (death) on plates without estradiol (Figure 5B). In addition, growth rate analysis showed that in YPD medium supplemented with 30 nM estradiol all three SPAZ SG

strains show growth similar to the BY4741 wild-type strain (Figure 5C). Metabolomics analyses showed no significant change in any of the metabolites examined compared to a WT strain, further underscoring the normal physiology of strains safeguarded by this mechanism (Figure 5D and Supplementary Table S3).

Reducing background of a yeast TET promoter

In addition to the GAL promoters and their derivatives, other well-studied switches are based on the TET system, previously used in yeast (29) and other organisms. We wished to construct a TET-based SG strain. Unfortunately, the current TET system available to us (29, 30) was too leaky and showed no significant reduction in growth without Doxycycline for all of our 44 essential genes SGs. Therefore, based on success with the SPAL and SPAZ promoters, we assembled a "SPET" promoter. This promoter is based on the TET promoter used in our lab (Figure 6A) with a URS1 sequence inserted downstream of the TATA box and upstream of the estimated transcription start site (Figure 6A).

First, we cloned either a single URS1 (SPET) or 3 repeats of the URS1 sequence (3xSPET) into the TET promoter. These promoters were then cloned into an acceptor vector for safeguard gene integration (pSIB499). For examining the function of the SPET promoter we cloned one of our candidate essential genes, *SEC4*, under the control of the TET, SPET or 3xSPET promoters and analyzed the expression with or without Doxycycline (Figure 6B), compared to a strain carrying a WT copy of *SEC4*. The 3xSPET promoter showed almost 4-fold reduction in growth without doxycycline (YPD) compared to both the SPET and the TET promoters. However, although very slight, all three SPET promoters show a growth defect compared to expression of the SG genes driven by their native promoters. Thus, with relatively simple tweaks, the TET promoter can also serve as a SG switch.

Screening a library of decoy molecules

One key feature of a specific application for safeguard strains, namely to guard proprietary elements of the strain e.g. in industrial applications, is the ability to prevent outside sources from propagating the strains without previous knowledge of the very specific medium requirements of a given SG strain. In addition to enabling growth in very low concentrations of the required compound the identity of that compound could be masked using an array of distinct but inert "decoy" molecules. A major requirement for such decoy molecules is that they should not by themselves have any deleterious effect on the growth or physiology of the yeast strain. Ideally this should also be true of any ligand actually used to activate the SG. With this in mind, we analyzed the transcriptome of yeast cells (BY4741) in the presence of 22 compounds (Table S4), at two concentrations each to evaluate their physiologic impact. They were separated into three groups based on the solvent used (Water, DMSO or Ethanol) and the transcriptome were compared to a solvent control. Analysis of the transcriptome of cells exposed to the compounds revealed that in many cases there is a strong effect on transcription (Figure 7 and supplementary Table S5). However, several compounds produce only minor changes (1-3 genes) or even no change at all (Figure 7 and supplementary Table S5). The latter

group contains the best candidates as ligands for future gene regulatory systems for safeguards.

Discussion

The growing need for biocontainment in industrial production settings, and in industry and academic labs, and more important, as a containment mechanism for potential field release of recombinant organisms, has led to an investment in designing safeguard strains. Efforts have been made both in bacteria (10-12, 15-17) and yeast (18). Interestingly, it has been shown that in order to achieve an extremely stable safeguard strain with as few escapers as possible there is a need to multiplex distinct strategies of biocontainment. To do so it is imperative to have a large array of possible essential genes that are good candidates to act in a safeguard system.

Here screening a library of 250 essential genes and narrowed this down to three best-acting genes (*FAS2*, *RPB11* and *SEC4*) under estradiol induction. In our screen we also selected the best acting promoter for each gene out of 6 possible candidates. We eventually selected for constructs that showed the lowest possible escape rate with the least measurable effect on growth compare to a control strain. This screen allowed us to achieve an escape rate of less than 10^{-7} with only a single essential gene; compared to 10^{-6} observed previously (18) for a single gene and closer to 10^{-9} that was observed for two histone genes. In a diploid context this escape frequency dropped by at least another order of magnitude. It is quite possible that these three genes are good candidates for the GAL/estradiol system but might very well not work as well in when controlled by other ligand activated promoters. However, similar workflows can easily be utilized with any other transcription system and might identify additional suitable candidates.

We have also shown that using promoter engineering we can decrease a deleterious genome-wide effect on endogenous GAL promoters using synthetically designed zinc-finger (26) and TET promoters, by improving promoter leakiness using a natural URS sequence (Figure 5). Although further optimization is required to adjust the appropriate dynamic range of the Tet system, its addition to the SG toolbox will help multiplex SG constructs and may reduce escape rates even further.

Analysis of escapers indicated that the use of the URS sequences also represents a weak spot of these promoters, as almost all revertants mapped to genes encoding Rpd3L complex subunits that repress transcription of genes carrying a URS1 sequence (27). Knowing this, with the fact that in all cases the mutation was recessive, future safeguards could either be constructed in a diploid background or supplemented with extra copies of the Rpd3L complex genes to reduce escape rates significantly.

Nevertheless, the best strategy to decrease the escape rate in safeguard strains will be to use multiplexed safeguards that will serve as backups to one another in case one fails due to naturally occurring mutations, which are part of any biosystem (18).

Tightly controlled genetic switches regulating essential genes are also useful in many other genetic contexts such as the evaluation of essential gene function (i.e. gene shutoff studies) and perhaps in the study of interactions of essential genes in genome wide contexts (31).

Material and Methods

Strains, plasmids, oligonucleotides and media

Yeast strains and the plasmids contained are listed in Supplementary Table S6. All SG strains are derived from BY4741 (*MATa leu2Δ0 met15Δ0 ura3Δ0 his3Δ1*) and BY4742 (*MATα leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1*) (22). Oligonucleotides used to amplify essential genes for shuffle plasmid construction are listed in Supplementary Table S7, other Oligonucleotides and primers used in this work are available upon request. Media used were as follows. Yeast strains were cultured in YPD medium, YP-galactose medium, or SD-based media supplemented with appropriate amino acids; fully supplemented medium containing all amino acids plus uracil and adenine is referred to as SC (32, 33). β-Estradiol was purchased from Sigma-Aldrich (St. Louis, MO), and 5-fluoroorotic acid (5-FOA) was from US Biological (Massachusetts, MA). Doxycycline (Dox) was obtained from Clontech laboratories (Mountain View, CA). Escherichia coli was grown in Luria Broth (LB) media. To select strains with drug-resistant genes, carbenicillin (Sigma-Aldrich) or kanamycin (Sigma-Aldrich) were used at final concentrations of 75 μg/ml and 50 μg/ml respectively. Agar was added to 2% for preparing solid media. Final concentrations of compounds and their solvents are listed in Table S4, 1000X stock solution was prepared for all compounds. Plasmids used are described in Table S8. SPAZ promoter sequence and SPET promoters sequence are listed in Supplementary data.

Plasmid recovery from yeast

Plasmid recovery from yeast was carried out using a Zymoprep yeast plasmid miniprep kit (Zymo Research, Irvine, CA) following the manufacturer's instructions.

Safeguard promoter identification

For each strain, genomic DNA was extracted and the promoter region was amplified using the vector forward primer and a reverse primer 100bp downstream of the ATG. Each PCR fragment was sent for sequencing using the same primers. The results of the sequencing are summarized in Table 1.

Escape rate measurement

Escape rates were calculated using the method of the median (34). For measurement rates in the 10^{-5} to 10^{-9} range, 5–12 independent cultures (each grown from a single parent colony) were inoculated into 20 mL of media supplemented with 30nM estradiol in liquid cultures. In total, 10^8 and 10^7 cells were plated on restrictive medium. Viable titer was determined by plating 100 μL of a five to six serial 10-fold dilution on permissive medium. The reversion frequency was obtained by dividing colony-forming units on

restrictive plates by colony-forming units on permissive plates. The median reversion frequency was then used to calculate the rate using the method of the median.

For each SG strain to be measured, 8 "escaper" escaper colonies were picked from independently grown cultures and grown up in 10 mL of permissive liquid culture (with plasmid selection if applicable), for 48 h at 30°C. Tenfold serial dilutions were plated on restrictive and permissive agar plates, and incubated at 30°C for 2–3 d until single colonies appeared. One colony was chosen per culture to assure independence.

Creating parts for combinatorial yeast Golden-Gate (yGG).
In addition to the 47 strains chosen from the TET-off library we decided to add three additional genes (*HSP10, RPC11* and *SUI1*) previously shown to be potential safeguard genes (18). All 50 genes were to be potentially cloned downstream of 7 different promoters: *GAL1, GAL10, GAL7, SPAL2, SPAL10, SPAL7* (24). They were assembled with the *GAL1* terminator and into an acceptor vectors with integration cassettes (23). All combinations were examined for their ability to serve as potential safeguards. All 50 genes, 6 promoters and terminator were amplified with the appropriate overhangs for yGG from the yeast genome (23), cloned into a pCR-Blunt II-TOPO vector (Invitrogen) and sequence verified. Most genes were amplified as one part genes, however, some were either too large to amplify as one part or had an internal *Bsa*I site, which is incompatible with yGG assembly (23), and thus were amplified as multiple gene parts compatible with yGG assembly. All but one gene, *PSA1*, failed at this step, so we moved forward with the successful 49 genes. In total we synthesized 67 gene parts, 6 promoter parts and one terminator part for the yGG assembly. As an acceptor vector we chose to use pAV10.HO5.loxP, an integrating vector, with the *LEU2* marker for integration into the *HO* locus (23).

DNA-Seq and RNA-Seq
Pair-end whole genome sequencing of escapers was performed using an IlluminaHiSeq2500 using TruSeq preparations kits. In total, 20 samples were sequenced with 19.5M – 33.5M paired reads generated per sample. The length of each read was 101 base pairs. Quality control was performed using software FastQC version 0.11.2 (35). All the reads in FASTQ format were aligned to the *S. cerevisiae* reference genome constructed starting with the sequence for control strains (strain BY4741 BY4742 genome sequences) using BWA version 0.7.8 with -P -M –R parameter settings (36). Approximately 86%-93% reads were aligned to the corresponding reference genome. Software SAMtools version 0.1.19 was used to call variants with mpileup -A -uf and bcftools view -bvcg parameter settings (37). The results were subjected to a set of post-processing filters requiring (i) a minimum of 10x coverage per variant site; (ii) wild-type reads in < 10% of the total reads per site; (iii) Reads supporting the variant of the control sample in < 5%. RNA-Seq datasets were generated and analyzed as described in (18).

Metabolite Extraction
Intracellular metabolites were extracted from yeast, using a method adapted from (38). Briefly, overnight cultures grown in the appropriate dropout SC medium with 30nM estradiol, and were then diluted to OD 0.1 and grown to OD 1.0 in the same medium.

12.5 mL of culture was then rapidly quenched by addition into 37.5 ml of 60% methanol/10 mM Tricine, pH 7.4, that was maintained at −40°C to stop metabolism. After 5 min at −40°C, cells were spun at 2,500 × g for 2 min at 0°C, washed with 1 ml of the same buffer, and then resuspended in 1 ml of 75% ethanol/0.5 mM Tricine, pH 7.4. Intracellular metabolites were extracted by incubating at 80°C for 3 min, followed by incubation at 4°C for 5 min. Samples were spun at 16,100 × g for 1 min to pellet cell debris, and 0.9 ml of the supernatant was transferred to a new tube. After a second spin at 16,100 × g for 1 min, 0.8 ml of the supernatant was transferred to a new tube, dried down using a speed-vac, and stored at −80°C until analysis.

LC-MS/MS analysis

Dried metabolites were analyzed as previously described by (39). The dried metabolites are resuspended in one of 3 buffers which allow a specific subset of metabolites to be retained by the LC column, which are then eluted off by a methanol gradient. This explains in part why the number of times a metabolite is listed is variable. Some metabolites can be detected by more than one of the three methods. Furthermore, for each method, there are usually 2 listings for each metabolite that represent the two most abundant daughter fragments. Occasionally, only one daughter fragment can be detected reliably, which is why a few metabolites are only listed once.

Acknowledgments

We thank Boeke lab members for helpful discussions. This work was supported by the Defense Advanced Research Project Agency [N66001-12-C4020 (to J.D.B.)].

References

1. Cohen SN, Chang AC, Boyer HW, & Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 70(11):3240-3244.
2. Ellis T, Adie T, & Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative biology : quantitative biosciences from nano to macro* 3(2):109-118.
3. Wang HH*, et al.* (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460(7257):894-898.
4. Isaacs FJ*, et al.* (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333(6040):348-353.
5. Hillson NJ, Rosengarten RD, & Keasling JD (2012) j5 DNA assembly design automation software. *ACS synthetic biology* 1(1):14-21.
6. Czar MJ, Cai Y, & Peccoud J (2009) Writing DNA with GenoCAD. *Nucleic acids research* 37(Web Server issue):W40-47.
7. Cai Y, Hartnett B, Gustafsson C, & Peccoud J (2007) A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. *Bioinformatics* 23(20):2760-2767.
8. Wright O, Stan GB, & Ellis T (2013) Building-in biosafety for synthetic biology. *Microbiology* 159(Pt 7):1221-1235.
9. Steidler L*, et al.* (2003) Biological containment of genetically modified Lactococcus lactis for intestinal delivery of human interleukin 10. *Nature biotechnology* 21(7):785-789.
10. Poulsen LK, Larsen NW, Molin S, & Andersson P (1989) A family of genes encoding a cell-killing function may be conserved in all gram-negative bacteria. *Molecular microbiology* 3(11):1463-1472.
11. Knudsen SM & Karlstrom OH (1991) Development of efficient suicide mechanisms for biological containment of bacteria. *Applied and environmental microbiology* 57(1):85-92.
12. Gerdes K*, et al.* (1986) Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon. *The EMBO journal* 5(8):2023-2029.
13. Bej AK, Perlin MH, & Atlas RM (1988) Model suicide vector for containment of genetically engineered microorganisms. *Applied and environmental microbiology* 54(10):2472-2477.
14. Wright O, Delmans M, Stan GB, & Ellis T (2015) GeneGuard: A modular plasmid system designed for biosafety. *ACS synthetic biology* 4(3):307-316.
15. Ronchel MC & Ramos JL (2001) Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Applied and environmental microbiology* 67(6):2649-2656.
16. Rovner AJ*, et al.* (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518(7537):89-93.
17. Mandell DJ*, et al.* (2015) Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518(7537):55-60.
18. Cai Y*, et al.* (2015) Intrinsic biocontainment: Multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes.

Proceedings of the National Academy of Sciences of the United States of America 112(6):1803-1808.
19. Mortimer RK (2000) Evolution and variation of the yeast (Saccharomyces) genome. *Genome research* 10(4):403-409.
20. Mnaimneh S*, et al.* (2004) Exploration of essential gene functions via titratable promoter alleles. *Cell* 118(1):31-44.
21. Boeke JD, Trueheart J, Natsoulis G, & Fink GR (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods in enzymology* 154:164-175.
22. Brachmann CB*, et al.* (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14(2):115-132.
23. Agmon N*, et al.* (2015) Yeast Golden Gate (yGG) for efficient assembly of S. cerevisiae transcription units. *ACS synthetic biology*.
24. Vidal M, Brachmann RK, Fattaey A, Harlow E, & Boeke JD (1996) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* 93(19):10315-10320.
25. McIsaac RS*, et al.* (2011) Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. *Molecular biology of the cell* 22(22):4447-4459.
26. McIsaac RS*, et al.* (2013) Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic acids research* 41(4):e57.
27. Keogh MC*, et al.* (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123(4):593-605.
28. Sumrada RA & Cooper TG (1987) Ubiquitous upstream repression sequences control activation of the inducible arginase gene in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 84(12):3997-4001.
29. Gari E, Piedrafita L, Aldea M, & Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast* 13(9):837-848.
30. Ikushima S, Zhao Y, & Boeke JD (2015) Development of a Tightly Controlled Off Switch for Saccharomyces cerevisiae Regulated by Camphor, a Low-Cost Natural Product. *G3 (Bethesda)* 5(10):1983-1990.
31. Magtanong L*, et al.* (2011) Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nature biotechnology* 29(6):505-511.
32. Kaiser C, Michaelis S, Mitchell A, & Cold Spring Harbor Laboratory. (1994) *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) 1994 Ed pp vii, 234 p.
33. Sherman F (2002) Getting started with yeast. *Methods in enzymology* 350:3-41.
34. Lea DE & Coulson CA (1949) The distribution of the numbers of mutants in bacterial populations. *Journal of genetics* 49(3):264-285.
35. Andrews S (FastQC a quality-control tool for high-throughput sequence data.
36. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.

37. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
38. Castrillo JI, Hayes A, Mohammed S, Gaskell SJ, & Oliver SG (2003) An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* 62(6):929-937.
39. Tu BP*, et al.* (2007) Cyclic changes in metabolic state during the life of a yeast cell. *Proceedings of the National Academy of Sciences of the United States of America* 104(43):16886-16891.

Table 1: SG strains promoters and escape rates.

Essential Gene	Promoter	Escape rate*
FAS2	*SPAL5*	6.6E-08
HTS1	*SPAL6*	7.0E-08
RPB11	*SPAL5*	1.9E-08
SEC17	*SPAL2*	8.8E-08
SEC4	*SPAL2*	6.5E-08

* Escape rates were calculated according to method of the median (34)

Table 2: Escaper complementation group analysis

GENE	Complementation Groups	R/D*
FAS2	Group 1: *MAT**a*** (R1, R4, R5, R8) and *MAT*□ (R9, R10, R13, R14, R15)	R
	Group 2: *MAT**a*** R2	R
	Group 3: *MAT **a*** R3	R
	Group 4: *MAT**a*** R6	R
	Group 5: *MAT**a*** R7	R
	Group 6: *MATα* R11	R
	Group 7: *MATα* R12	R
	Group 8: *MATα* R16	R
RPB11	R3	D
	All others same complementation group	R
SEC4	Group 1: *MAT**a*** (R2, R4, R5, R6, R7) and *MAT*□ (R11, R12, R13, R15, R16)	R
	Group 2: *MAT**a*** R1	R
	Group 3: *MAT**a*** R3	R
	Group 4: *MAT**a*** R8	R
	Group 5: *MATα* R9	R
	Group 6: *MATα* R10	R
	Group 7: *MATα* R15	R
		R

* R=recessive, D=Dominant;

Table 3: Genomic mutations found in escapers by DNAseq.

SG	Escaper	Gene	Mutation
FAS2	R1	*UME6*	E75*
FAS2	R2	*RPD3*	M1V
FAS2	R3	*RPD3*	G312D
FAS2	R6	*DEP1*	Y236*
FAS2	R7	*YTA12*	T971A[&]
		SDS3	E71*
FAS2	R11	*RHO1*	D129N[&]
		SIN3	3061insT†
FAS2	R12	*SIN3*	Y630*
FAS2	R16	*DEP1*	Y242*
RPB11	R9	*UME6*	C165*
SEC4	R1	*RPD3*	G309C
SEC4	R2	*UBP13*	L67M[&]
		UME6	Q113*
SEC4	R3	*SDS3*	Q65*
SEC4	R9	*SRL2*	K74L[&]
		DEP1	E222*
SEC4	R10	*DEP1*	512insC†
SEC4	R14	*UME6*	W368*

*** indicates premature stop codon; † indicates frameshift; [&] indicates assumed secondary mutation. Underlined genes encode subunits of the RPD3L deacetylase.**

Figure and Figure legends

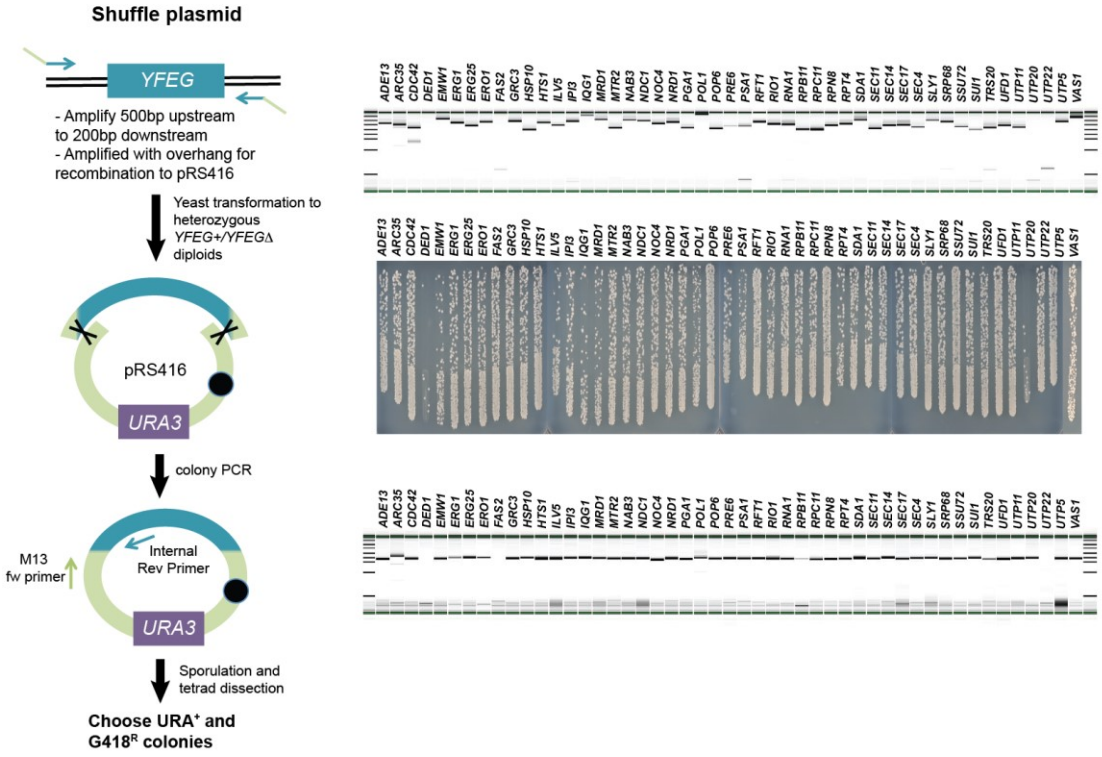

Figure 1: Outline of steps to create shuffle strains. Genes were amplified from the yeast genome with 500 bp upstream and 200 bp downstream as well as 50 bp in each end for recombination with the pRS416 vector. *Sma*I digested pRS416 and each amplified essential gene were transformed into the appropriate heterozygous diploid strain. Transformations (30µL) were dripped onto Omnitrays and single colonies were picked. Shuffle plasmid cloning was verified using plasmid primers and a gene internal primer. Following plasmid verification all strains were sporulated and G418 resistant Ura^+ colonies were saved (both *MAT**a*** and *MATalpha* strains were saved). YFEG, Your favorite essential gene.

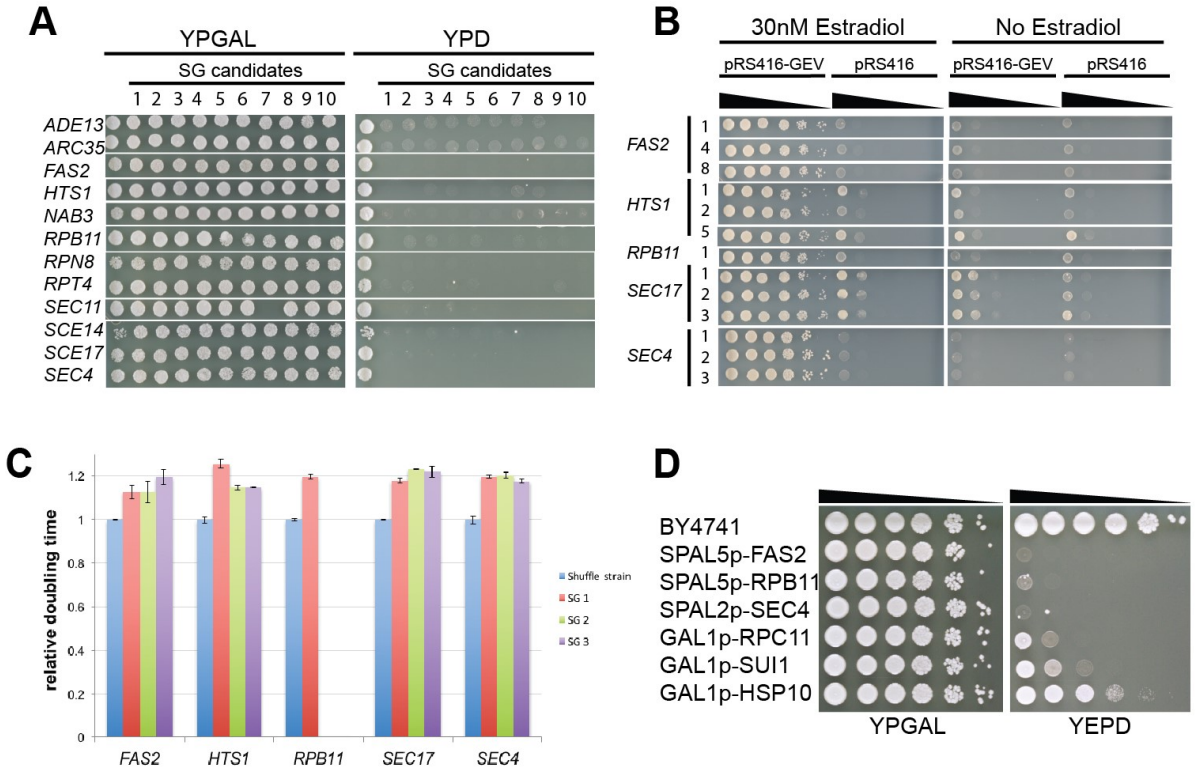

Figure 2: Screening for best safeguard strain after integration of the SG construct. Following transformation into the shuffle strains, and plating on 5-Foa, 10 candidates from each essential gene were examined by plating on YPGAL vs. YPD. (A) 12 best performing strains, were chosen for GEV plasmid transformation. 8 isolates from each SG strain were transformed with GEV containing plasmid and analyzed using a dot assay on Dextrose without Estradiol. (B) 13 good candidates that grew well on SC–Ura without estradiol and did not grow on SC–Ura without estradiol were chosen for further analysis. (C) Liquid cultures were diluted and subjected to OD measurements every 10 min for 24 h. A growth curve was created and doubling time was calculated for each strain. The experiment included 3 independent cultures for each strain in order to calculate SD. (D) A comparison of our three best SG strains with the previously published ones (18) shows a significantly lower growth on YPD medium compared to YPGal, indicating that this screen identified better performing SG strains.

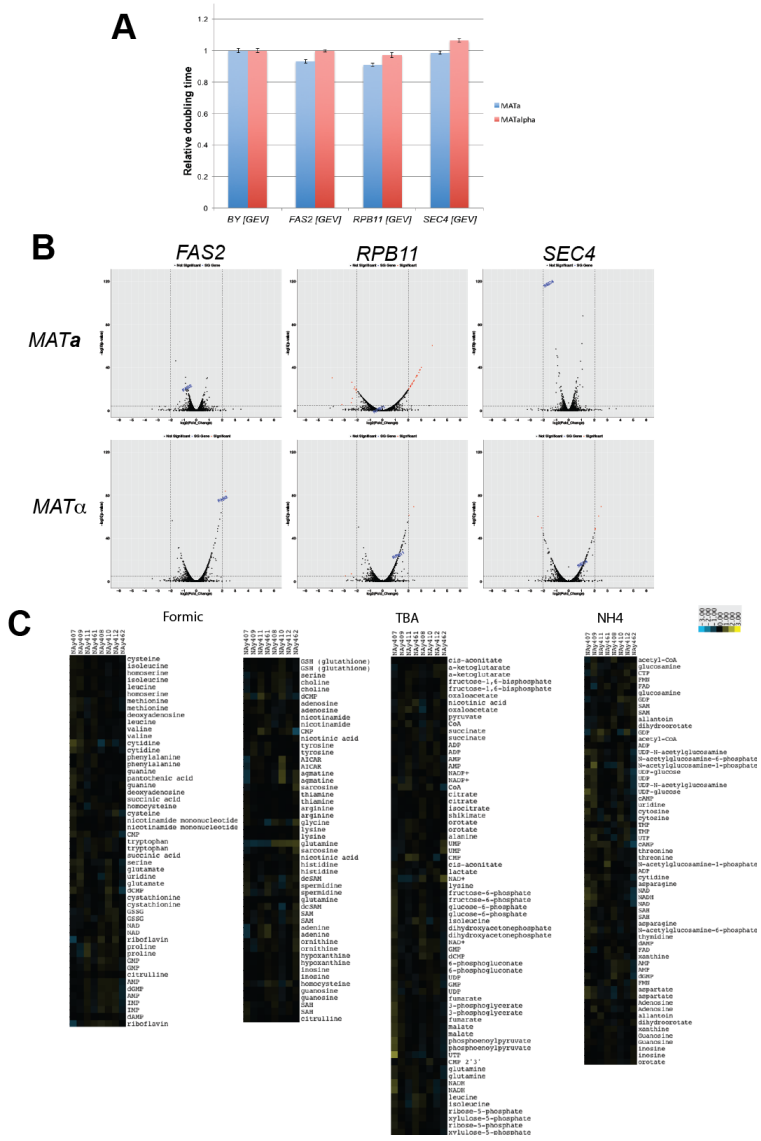

Figure 3: *MATa* and *MATα* safeguard strains compared to WT strain with GEV plasmid. (A) Graphic representation of relative doubling time of SG strains compare to WT with GEV plasmid. Liquid cultures were diluted and subjected to OD measurements every 10 min for a 24 h period. Doubling time was calculated as in Fig. 2. (B) Transcriptome profiling of the various safeguard strains. The graph is organized by gene/promoter pairs. Red dots in the volcano plots represent statistically significantly dysregulated genes (see Table S1 for lists of genes affected). Blue labeled dot represents SG gene in each sample. The transcriptome profiling shows limited transcriptome changes to the safeguard strains compared with the wild type. (C) Metabolomics analysis presented as a heat-map of the fold change for each metabolite analyzed. Formic (formic acid 0.1%) , TBA (10mM pH 5.0) and NH4 (5mM ammonium acetate) represents the solvents used to extract the yeast. *MAT**a*** Safeguard strains: NAy407 (*FAS2*), NAy409 (*RPB11*) and NAy411 (*SEC4*), *MAT**a*** control – NAy461. *MATα* Safeguard strains: NAy408 (*FAS2*), NAy410 (*RPB11*) and NAy412 (*SEC4*), *MATα* control – NAy462.

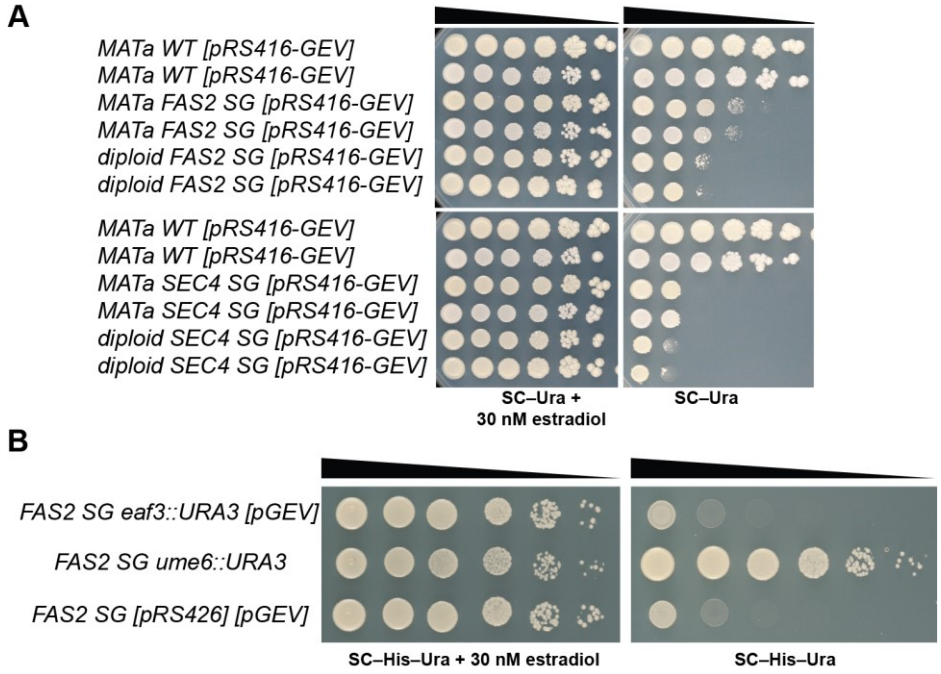

Figure 4: Analysis of safeguard escapers. (A) We hypothesized that because all of *FAS2* and *SEC4* escapers were due to recessive mutations, diploid safeguard strains will have lower escape rates. Serial dilutions of *FAS2* and *SEC4* diploid safeguard strain were compared to the original haploid safeguard strains and wt strains on the appropriate medium with and without estradiol. This shows that indeed the diploids show decrease escape rate compared to the haploid strains. (B) Genome sequencing revealed that all escapers carry a mutation in genes encoding subunits of the Rpd3L complex. In order to reconstitute the escaper phenotype in our SG strains we deleted *ume6* in the *FAS2 MAT**a*** SG strain. Serial dilution of *FAS2* SG deleted for *ume6, eaf3* and carrying a pRS416-GEV plasmid were plated in appropriate medium with and without estradiol. Cell deleted for *ume6* completely recapitulated the escapers phenotype.

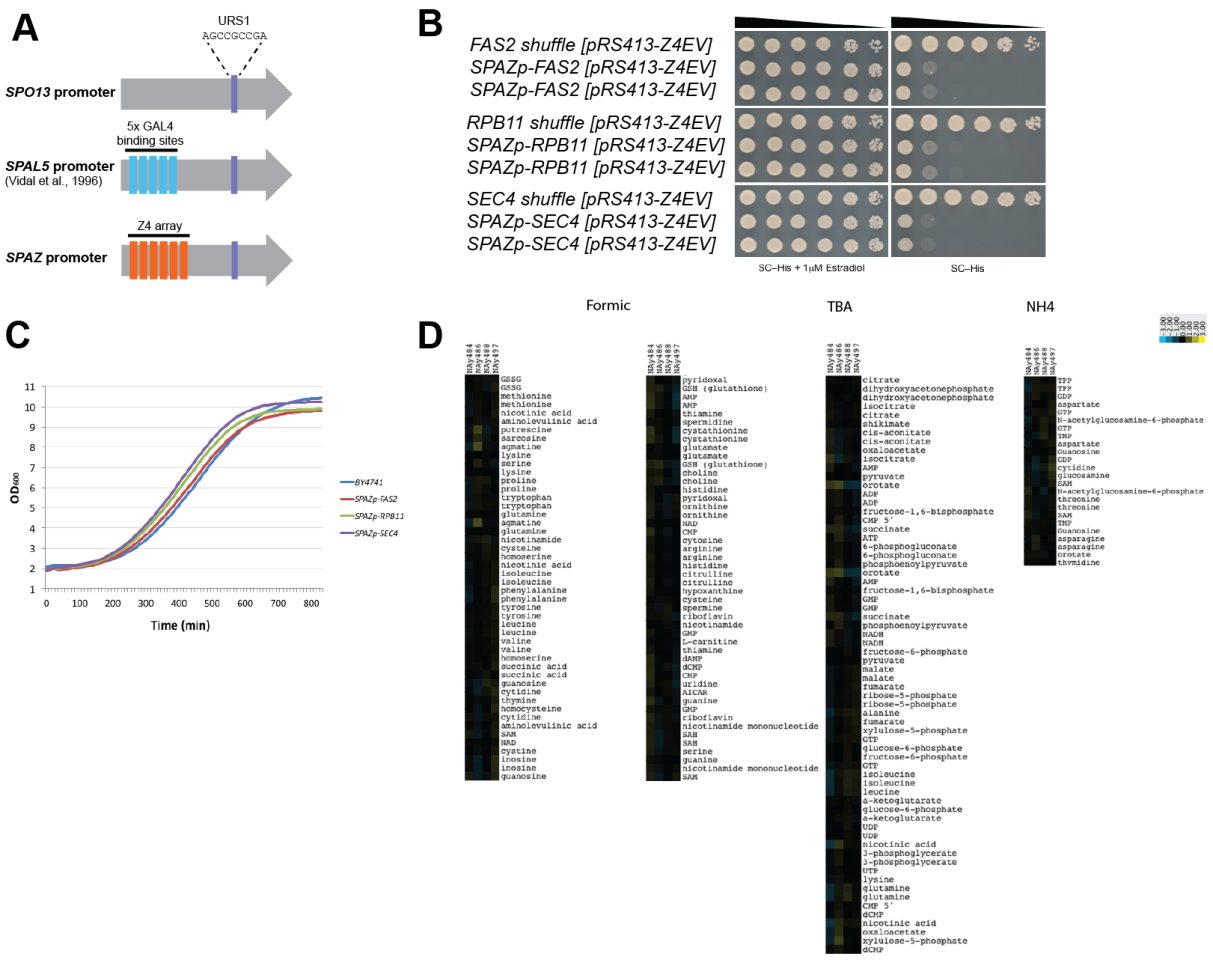

Figure 5: The *SPAZ* promoter. (A) Schematic representation of the *SPO13* promoter with URS1 sequence (purple box), the *SPAL5* promoter with 5 GAL4 binding sites (turquoise boxes) and the *SPAZ* promoter with the 6 repeats of the Z4 binding site (orange boxes). (B) The *SPAZ* promoter used to drive our 3 top safeguard genes: *FAS2, RPB11* and *SEC4*. Serial dilutions of all strains compared to their shuffle strains containing the ZEV expressing plasmid were plated on YPD media with or without 1 μM estradiol. (C) Growth of all three SPAZ safeguard strains was measured compared to BY4741 wild type strain in YPD medium containing 30 nM estradiol. (D) Metabolomics analysis presented as a heat-map of the fold change from each metabolite analyzed. Safeguard strains used were NAy484 (*FAS2*), NAy486 (*RPB11*) and NAy488 (*SEC4*), control – NAy497.

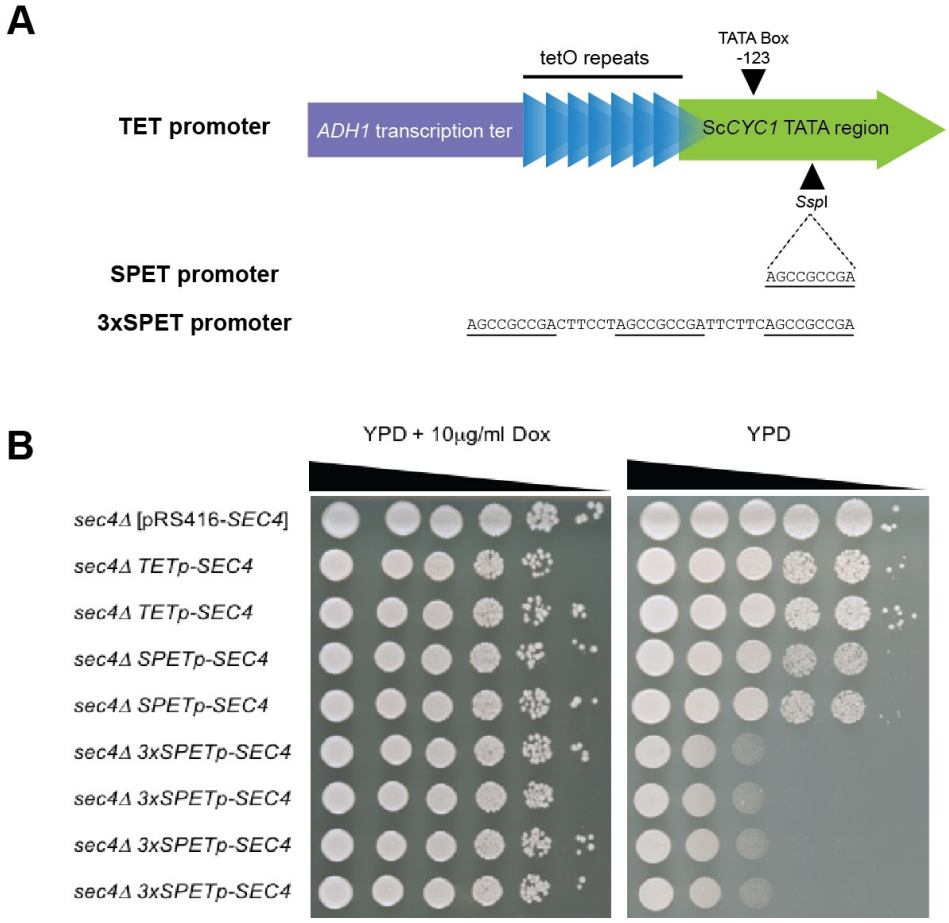

Figure 6: Engineering the TET promoter (A) Schematic representation of the *TET*, *SPET* and *3xSPET* with ADH1 transcription terminator sequence (purple box), the *tetO* repeats (blue triangles) and *CYC1* TATA region (green arrow), arrow heads indicate TATA box location and *SspI* restriction site. For the *SPET* promoter a single URS1 sequence was cloned into the *SspI* site. for *3xSPET* three repeats of the URS1 sequence (with linkers) were cloned to the *SspI* site. (B) The *TET* (2 isolates), *SPET* (2 isolates) and *3xSPET* (4 isolates) promoters used to drive the *SEC4* gene. Serial dilutions of all strains compared to the shuffle strain were plated on YPD medium with or without 10μg/ml Doxycycline.

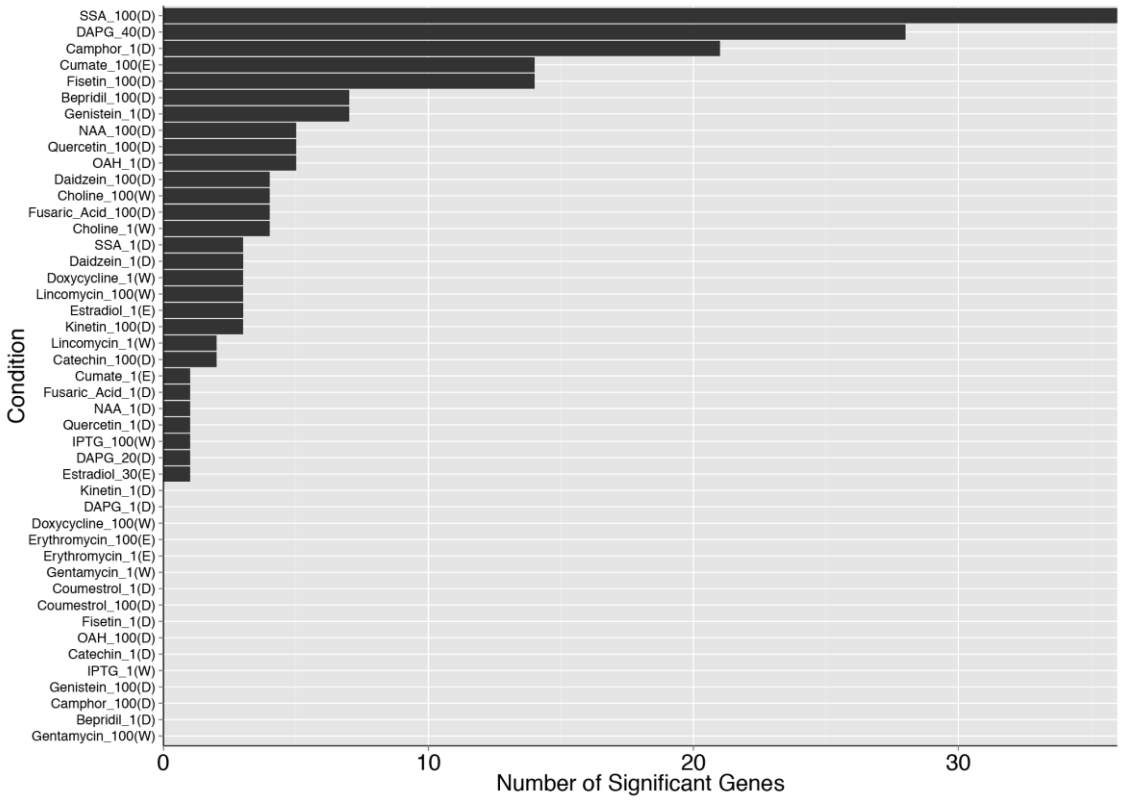

Figure 7: **Decoy molecules - effect on transcription**. Graphical representation of the number of transcripts/genes changed in response to each specific candidate decoy compound and concentration. BY4741 WT cells were grown in liquid culture in the presence of each candidate decoy molecule (_1 and _100 represent 1 and 100 μM respectively) and subjected to transcriptome analysis. Compounds/concentrations useful in "decoy mode" are indicated in red.