Natural Genetic Variation Can Independently Tune the Induced Fraction and Induction Level of a Bimodal Signaling Response

Jue Wang[1,2,3], Julius Palme[4], Kayla B. Lee[2,5], Michael Springer[2,*]

[1]Systems Biology Graduate Program, Harvard University, Cambridge, Massachusetts 02138

[2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts.

[3]Ginkgo Bioworks, 25-27 Drydock Avenue 8th Floor, Boston, MA 02210

[4]Plant Systems Biology, School of Life Sciences Weihenstephan, Technische Universität, München, 85354 Freising, Germany

[5]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts.

*To whom correspondence should be addressed: michael_springer@hms.harvard.edu

Abstract

Bimodal gene expression by genetically identical cells is a pervasive feature of signaling networks, but the mechanisms modulating bimodality are poorly understood. We found that natural yeast strains induce the galactose-utilization (GAL) pathway with a variety of bimodal phenotypes in mixtures of glucose and galactose. The phenotypic variation can be described in terms of two uncorrelated features representing the fraction of cells that are induced and the expression level of the induced subpopulation. We mapped genomic loci underlying these two traits using bulk-segregant analysis, identified causal genes in 3 loci, and phenotyped allele-replacement strains containing all allelic combinations of these genes. One gene affected only the induced fraction of the GAL response, another affected only the level of induction, and a third gene affected both traits. Additionally, the genetic effect on induced fraction could be phenocopied by varying the growth conditions prior to galactose induction. Our results show that different quantitative features of a bimodal signaling response can be tuned independently by genetic and environmental perturbations, and that this tuning can change the response from unimodal to bimodal. This modularity may help cells adapt to complex natural environments on physiological as well as evolutionary timescales.

Introduction

Non-genetic heterogeneity is a pervasive feature of gene expression and cellular signaling [1–3]. Bimodal responses, where cells in an isogenic population adopt one of two distinct states, are particularly important for microbes coping with fluctuating environments [4,5] and cells of multicellular organisms differentiating into discrete types [6,7]. The galactose-utilization (GAL) pathway in *Saccharomyces cerevisiae* (budding yeast) is a well-characterized bimodal response and a classic model of microbial decision-making [8,9]. GAL enzymes are tightly repressed in glucose and activated almost 1000-fold in galactose [10]. In mixtures of glucose and galactose, GAL genes induce as a function of the galactose-to-glucose ratio [11] and display complex patterns of bimodal expression [12].

Bimodality of GAL gene expression is attributed to bistability arising from positive feedback through the Gal1p kinase and Gal3p transducer [13,14]. However, perturbing other pathway components such as Gal2p permease, Gal4p activator, and Gal80p repressor also affect quantitative features of the GAL response [14–17]. Additionally, the modality of the GAL response is affected by the metabolic conditions prior to encountering galactose [12]. Despite the complex response of GAL expression distributions to genetic and environmental perturbations, most studies of the pathway have focused on one quantitative feature such as the induced fraction [16,18,19], with a few recent exceptions [13,20]. How multiple quantitative features of the pathway are controlled and vary across perturbations is poorly understood.

In previous work, we found that natural yeast isolates differed widely in the inducibility of GAL genes in glucose + galactose mixtures [19,21]. In particular, some strains displayed bimodal activation of GAL genes while other strains were unimodal in the same conditions. Similar

population heterogeneity has been seen in yeast maltose utilization [22] and bacterial utilization of various sugar mixtures [23]. This natural variation provides an opportunity to dissect the genetic variants modulating bimodality in nature and expand our knowledge of the repertoire of quantitative behaviors that can be achieved by this model circuit.

In this work, we showed that natural yeast isolates induce the GAL pathway with a diverse array of bimodal and unimodal expression patterns that vary with sugar conditions. We analyzed this variation in terms of two traits representing the induced fraction of cells and the expression level the induced subpopulation, which vary in an uncorrelated way across natural isolates. Using bulk segregant analysis and CRISPR/cas9 allele replacement, we identified genetic variants underlying these two traits and showed that the variants can affect the traits independently of each other. Additionally, we found that the metabolic history of cells before inducing GAL genes also affects the bimodal response in a trait-specific way. The independent tuning of these two quantitative features of the GAL response can account for the diversity of unimodal and bimodal phenotypes observed in our natural isolates. This genetic flexibility may be advantageous for cells adapting to complex natural nutrient environments.

Results

Natural yeast isolates vary in the degree of bimodality of GAL induction

To study natural variation in the population behavior of the GAL pathway response, we measured the expression of a GAL1 promoter driving YFP (GAL1pr-YFP) in 34 geographically and ecologically diverse yeast strains [21,24,25] grown in a titration of glucose plus a constant level of galactose. As expected, we found that all strains are uninduced in high glucose and fully induced in low or no glucose (Figure 1). However, at intermediate glucose concentrations, some

strains display a unimodal population with intermediate (i.e. sub-maximal) GAL expression (Figure 1B) while other strains display a bimodal mixture of uninduced and partially induced cells (Figure 1C). Additionally, strains with the same modality still have quantitatively different GAL induction profiles (Figure S1), raising the question of what mechanisms can give rise to these diverse signaling phenotypes.

Variation in GAL bimodality phenotypes can be parameterized by two uncorrelated metrics

Upon close inspection, the GAL induction phenotypes generally seem to be a mixture of two components: an induced subpopulation that decreases in YFP level as glucose increases, and an uninduced subpopulation that remains at the same YFP level regardless of glucose concentration (Figure S2A). The mixing of these components can be quantified as an induced fraction that decreases as a function of glucose concentration (Figure S2B). Simply by varying the glucose-dependence of the induced fraction and of the induced subpopulation expression level, we can simulate many bimodal phenotypes, as well as unimodal phenotypes, reminiscent of the observed data (Figure S1, S4.2B-C). In this framework, a strain which is unimodal in a particular condition has an induced fraction of one (but a sub-maximal induced level), while a strain that is bimodal in this condition has an induced fraction of less than one.

Applying this population decomposition framework to our data, we computationally separated induced and uninduced cells from each GAL reporter distribution (Figure 2E) and calculated two summary metrics for each strain's phenotype: E_{10}, the glucose concentration where the induced subpopulation reaches 10% of its maximal GAL expression level, and F_{50}, the glucose concentration where 50% of cells in the population are induced (Figure 2F). For convenience, these metrics are in log2-transformed units, so a strain with $E_{10} = -1$ has an induced

subpopulation that reaches 10% of maximal induction at $2^{-1} = 0.5\%$ w/v glucose. We find that the E_{10} and F_{50} are uncorrelated across natural isolates, suggesting the possible existence of genetic changes that can decouple them (Figure 2G).

Bulk segregant analysis identifies genetic loci associated with GAL induction variation

To analyze the genetic basis of E_{10} and F_{50}, we crossed strains S288C and DBVPG1106 and phenotyped random haploid segregants from their hybrid. These parent strains differ in both traits, and their segregants display semi-continuous, correlated variation in these traits with a small number of outliers. Therefore, E_{10} and F_{50} are likely modulated by multiple genes, at least some of which affect both traits.

To identify these genes, we performed bulk-segregant linkage mapping using a pooled sorting strategy (Figure 2B). We chose a glucose+galactose condition where the parental GAL1pr-YFP distributions were maximally different and used it to induce a pooled mixture of haploid (MATa) segregants (Figure 2C). We then used FACS to sort the segregants into pools of uninduced ("OFF"), induced and low-expression ("LOW"), and induced and high-expression ("HI") cells (Figure 2C), and sequenced each pool to 15-33x coverage. We expected that a genomic locus affecting the induced level (and thus E_{10}) will differ in allele frequency between the LOW and HI pools, while any locus affecting the induced fraction (and thus F_{50}) would differ in parental allele frequency between the OFF pool and a computationally merged LOW+HI pool ("ON") (Materials and Methods).

We found 5 loci with significantly different allele frequencies between OFF/ON pools or between LOW/HI pools, defined as genomic regions with a peak log-odds (LOD) score > 5 calculated by MULTIPOOL [26] (Figure 3D; Materials and Methods). To look for causal

variants, we inspected gene annotations in a region of 2-LOD decrease around each LOD peak. The three most significant loci are centered at chrIV:457Kb, chrXIV:457Kb, and chrXVI:81Kb and contain the genes *GAL3*, *MKT1*, and *GAL4*, respectively (Figure 3D). *GAL3* and *GAL4* are direct regulators of galactose sensing, while *MKT1* is known to have pleiotropic effects in crosses between S288C and natural isolates . The *GAL3*-associated locus was significant only in the OFF/ON comparison, while the *GAL4*-associated locus was only significant in the LOW/HI comparison, suggesting that the effect of these loci are specific to either E_{10} or F_{50}. The *MKT1*-associated locus was significant in both comparisons but had a higher LOD score in the LOW/HI than in the OFF/ON comparison. Unlike the other loci, the *GAL4*-associated locus was enriched for the S288C allele in the DBVPG1106-like segregant pool, suggesting the possibility of transgressive segregation. Two other loci, at chrXII:1053Kb and chrXIII:105Kb, were also significant and seemed to have phenotype-specific effects, but did not contain any obvious genes for follow-up. Therefore, we focused on the chrIV (*GAL3*-associated), chrXIV (*MKT1*-associated), and chrXVI (*GAL4*-associated) loci for further investigation.

GAL3 and _GAL4_ alleles specifically affect F_{50} and E_{10} while _MKT1_ alleles affect both traits

To test if *GAL3*, *MKT1*, and *GAL4* alleles are causal variants in the chrIV, chrXIV, and chrXVI loci, we used CRISPR/cas9 to replace the coding region and flanking regions of each gene (Materials and Methods) in both DBVPG1106 and S288C with the allele from the other parent (Figure S4). In DBVPG1106, replacing the endogenous *GAL3* allele with $GAL3^{S288C}$ shifted F_{50} in the direction of the S288C parent (Figure S4A). *MKT1* replacement also shifted F_{50} and had a small effect on E_{10} as well. *GAL4* replacement had a small but clear effect on E_{10} and no detectable effect on F_{50}. In the S288C background, allele replacements had similar trait-specificity but much smaller effects (Figure S4B-C). In both parental backgrounds, the *GAL4*

allele replacement resulted in a change in E_{10} away from the value of the other parent. This is consistent with the sign of allele-frequency differences of the *GAL4*-containing locus between LOW and HI pools (Figure 2D) and confirms that *GAL4* is a transgressive allele in this cross. Overall, these results show that *GAL3*, *MKT1*, and *GAL4* are causal variants in their respective loci and corroborate the allelic effects inferred from our bulk segregant analysis.

The single allele replacements only modestly altered the phenotype of the parent strains, suggesting that other genes make substantial contributions to the total phenotypic difference. Alternatively, there may be genetic interactions between our mapped genes such that allele replacement of 2 or 3 of them is sufficient to achieve conversion of one parental phenotype to the other. To assess these possibilities, we constructed all 16 combinations of strain background, *GAL3* allele, *MKT1* allele, and *GAL4* allele from either the DBVPG1106 or S288C parent, and measured E_{10} and F_{50} of 2 independent isolates of each of the 16 genotypes. We examined the resulting phenotypic landscape (Figure 3A) in terms of pairs of strains differing in the allelic status of one gene (or strain background) while other genetic factors are held constant. The effect of switching from the the DBVPG1106 genetic variant to the S288C variant can be visualized as a vector in E_{10} versus F_{50} space (Figure 3B-E) or as a trait difference (Figure 3F).

This analysis reveals that the trait-specificity of single genetic changes are broadly consistent across different genotypic backgrounds (i.e. combinations of strain background and alleles at the other loci). This can be seen in the fact that effect vectors from DBVPG1106 to S288C variants in E_{10}-F_{50} space are parallel (Figure 3B-E), or equivalently, that differential effects cluster by angle to the origin (Figure 3F). Across the combinatorial allele replacement strains, it is clear that *GAL3* allele predominantly affects F_{50}, *GAL4* mostly affects E_{10}, and *MKT1* affects both traits. For example, a strain with $GAL4^{S288C}$ has a lower E_{10} than the congenic strain with

152 *GAL4*^{DBVPG1106} for all such strain pairs. These results show that E_{10} and F_{50} can be tuned
153 independently in this cross.

154 In contrast to trait-specificity, the effect sizes of single genetic changes across the combinatorial
155 landscape are more complicated. In general, *GAL3* allele effects on F_{50} are large, spanning up to
156 half the phenotypic distance between the parents. *MKT1* allele effects on F_{50} are almost as great,
157 and combined with *GAL3* allele replacement, can essentially phenoconvert DBVPG1106 to
158 S288C, but only along the F_{50} axis (DSSD versus in DDDD in Figure 3A). However, the
159 reciprocal replacement in the S288C background has a more modest effect (SDDS versus SSSS
160 in Figure 3A). Consistent with these findings, the strain background effect on F_{50} (which can be
161 interpreted as the residual variation after allele replacements) varies widely, from negligible to
162 almost as large as that of *GAL3* or *MKT1*. For the E_{10} trait, *GAL4* allele effects span between a
163 third and half the phenotypic distance between the parents, but in the opposite direction required
164 for phenoconversion. Therefore, triple allele replacement strains DSSS (DBVPG1106 *GAL3*^{S288C}
165 *MKT1*^{S288C} *GAL4*^{S288C}) and SSSD still differ substantially in E_{10} from their respective wildtype
166 SSSS and DDDD strains. Overall, strain background has effects on both E_{10} and F_{50} in most
167 genotype backgrounds, indicating substantial variation in both traits not accessed by our allele
168 swaps.

169 **Genetic and environmental perturbations that affect F_{50} do not affect E_{10}**

170 Our results above show that F_{50} can be tuned independently of E_{10} by some genetic variants in
171 the S288C x DBVPG1106 cross. To see if this is true over a larger range of F_{50}, we analyzed
172 phenotypic data on S288C, BC187, and DBVPG1106 strains whose *GAL3* loci have been
173 replaced with a panel of natural *GAL3* alleles that we previously showed to underlie a spectrum

of GAL inducibility phenotypes [19]. As expected, F_{50} varied widely as the *GAL3* allele is changed (Figure 4A). However, variation in E_{10} with *GAL3* allele was minimal and driven almost entirely by the strain background (Figure 4B).

Previously, a laboratory yeast strain was found to induce GAL genes bimodally or unimodally depending on the carbon source prior to galactose induction [12]. To see how metabolic memory affects E_{10} and F_{50} across our natural isolates, we pre-grew six strains in raffinose, acetate, or glycerol prior to induction in glucose + galactose. These carbon sources are neither inducers of GAL genes nor signals for glucose catabolite repression [27]. Nevertheless, they caused the yeast strains to exhibit different F_{50} upon subsequent induction in glucose + galactose (Figure 4C). We did not observe a carbon source pre-conditioning effect on E_{10} (Figure 4D). This parallels the effect of *GAL3* alleles, and suggests that the independent tuning of E_{10} and F_{50} is a consequence of how the GAL circuit is integrated with carbon metabolism more broadly.

Independent tuning of F_{50} and E_{10} modulates the modality of the GAL response

The definitions of the E_{10} and F_{50} metrics (Figure 1F, S4.2) imply that tuning either parameter independently should alter the apparent number of modes in GAL expression distributions. Since F_{50} varies over a wider range of glucose concentrations than E_{10} does under the perturbations we tested, we asked if independently tuning F_{50} affects modality. Indeed, plotting GAL reporter distributions shows that a number of allele replacements are able to convert strains from being bimodal to unimodal and vice versa. For example, DBVPG1106 is bimodal, but replacing alleles with $GAL3^{S288C}$ and $MKT1^{S288C}$ increases its F_{50} and makes it unimodal (Figure 5A-B). Conversely, BC187 is unimodal, but decreasing its F_{50} by introducing $GAL3^{YJM978}$ makes it bimodal (Figure 5C-D). Finally, Y12-WashU, one of the most obviously bimodal strains, is

196 rendered unimodal when pre-conditioned in acetate rather than raffinose before galactose
197 induction (Figure 5E-F).

198 ## Discussion

199 ### Independent tuning and the molecular mechanism of bimodality

200 It is known that positive feedback on GAL gene expression through *GAL3* tunes the switching
201 rate of cells between uninduced and induced states [14] and is a key contributor to the bistability
202 of the pathway [13]. Changes in *GAL3* dosage affects the induced fraction of GAL genes [16],
203 and a panel of natural *GAL3* variants confers a spectrum of GAL induction phenotypes [19]. We
204 put these previous observations in context by showing that natural *GAL3* alleles specifically
205 affect the sugar threshold where individual cells to switch to an induced state, while the level of
206 induction in that state is set by *GAL4* and other unknown genes. Both these features combine to
207 yield the population level behavior of the circuit, including apparent patterns of bimodality.
208 Underscoring this point, we found that *GAL3* allele replacement is sufficient to convert a
209 unimodal response to bimodal, and vice versa, while the level of the induced subpopulation
210 remains unchanged. This degree of modularity in the quantitative behavior of the GAL circuit
211 was previously unappreciated.

212 *GAL4* is the transcription factor activating all inducible GAL genes [10,28]. Previously, changes
213 in dosage of *GAL4* was found to have no effect on the GAL induced fraction [16]. The S288C
214 variant of *GAL4* contains a non-conservative R95G mutation, as well as a conservative R508K
215 mutation, relative to DBVPG1106 and other natural isolates. Residue 95 is on a loop linking the
216 DNA-binding and regulatory domains of *GAL4* and directly participates in interactions with
217 Gal11p [29,30], a component of the RNA polymerase II mediator complex that enhances

expression of *GAL* genes [31]. These observations suggest that the S288C and DBVPG1106 *GAL4* alleles might differ in their ability to activate transcription of GAL genes. This effect could be specific to induced level if differences in *GAL4* activity only affect GAL promoters that are in an active state and the latter variable is separately dictated by feedback loops such as *GAL3*. An important question for future work is whether this scenario is quantitatively plausible in a mathematical model of the GAL circuit, and what general features of this and other circuits allow for independent tuning.

Modularity of the GAL pathway, genetic background, and metabolic state

We also find that *MKT1* alleles affect the GAL response and can play almost as large a role as *GAL3*. *MKT1* is involved in maintaining killer toxin [32], regulating translation [33], and affects numerous traits in crosses between S288C and natural isolates [34–39]. The S288C allele of *MKT1* is a loss-of-function variant relative to natural alleles and causes lower expression of mitochondrial genes [40,41]. In turn, deletion mutants of mitochondrial genes tend to exhibit aberrant GAL induction; this effect is more pronounced on the induced fraction than on induced level [42], echoing our observations. Therefore, it is likely that the effect of *MKT1* allele on GAL induction is due to perturbations to mitochondrial function.

We found that much of the variation in E_{10}, and to a lesser extent F_{50}, must be attributed to unknown alleles in the genetic background. This dovetails with other recent reports that many traits in yeast are dominated by large effects from one or a few loci but can be tuned quantitatively by many small-effect loci [19,42–44]. Moreover, *MKT1* is not a member of the canonical GAL pathway, and nor are any genes in 2 other loci that reached significance in our linkage mapping. Combined with observations that deletion mutants of up to quarter of all yeast

genes have quantitatively perturbed GAL signaling [42], our results indicate that decision-making circuits are not as modular with respect to genetic variation as is often assumed.

In addition to genetic effects on GAL induction, we found that culturing cells in raffinose, glycerol, or acetate prior to induction in a glucose + galactose led to very different GAL phenotypes. Raffinose is commonly used to pre-culture cells for GAL induction studies (including most experiments here) because unlike glucose, it does not visibly repress GAL genes [14,45]. We chose glycerol and acetate by the same criterion. Otherwise, however, these carbon sources elicit very different physiological responses. Raffinose is hydrolyzed to release fructose [46], which can then be fermented [47,48]. Glycerol and acetate, by contrast, must be utilized via respiration [49], which entails expression changes in many genes [50] as well as differences in ATP/ADP ratio and redox state [51]. Therefore, our results suggest that factors other than canonical glucose catabolite repression may be important in setting the inducibility of GAL genes.

Our results indicate that memory of metabolic state is encoded by the GAL circuit and persists even after the cells have reached steady-state in inducing conditions (Figure 5B, Materials and Methods). This appears to be a distinct phenomenon from the “memory” of glucose or galactose pre-induction conditions previously attributed to bistability of the GAL network [14,52]. However, the fact that pre-induction carbon source specifically affects F_{50}, just as *GAL3* allele does, suggests that this positive feedback loop may be a nexus of regulation of GAL genes by multiple signals in the cell. Indeed, recently it was shown that NAD(P) can directly bind Gal80p and thereby impact downstream GAL pathway expression [53,54]. Since these studies relied on bulk measurements, it will be interesting to revisit these investigations using quantitative, single-cell readouts of pathway behavior.

Physiological and ecological role of independent tuning

Our results raise the question of why independent tuning of induced fraction and induced level would exist in nature. Previously we showed that natural variation in the timing of GAL induction during diauxic growth leads to a fitness tradeoff—some strains prepare for glucose exhaustion at an upfront cost while others maximize their growth rate on glucose but suffer a diauxic lag [21]. Related work showed that both strategies could be implemented by the same strain as part of a bimodal response [20], and that this may be an evolutionarily stable strategy [55]. Under this framework, tuning E_{10} and F_{50} separately would allow the timing of the inducing population, and its level of induction, to evolve separately. This could provide fitness benefits in certain conditions, although exactly what these conditions are would depend on the quantitative details of the costs and benefits of induction, an interesting issue to be explored in future work.

Materials and Methods

Strains and media

Strains were obtained as described in [21]. An initial set of 42 strains were assayed in glucose gradients + galactose. Strains CLIB324, L-1528, M22, W303, YIIC17-E5, YJM975, YJM981 were excluded from downstream analysis due to poor growth in our media conditions. Strain 378604X was also excluded due to a high basal expression phenotype that was an outlier in our collection. The genetic basis of this behavior is likely an interesting topic for follow-up studies.

All experiments were performed in synthetic minimal medium, which contains 1.7g/L Yeast Nitrogen Base (YNB) (BD Difco) and 5g/L ammonium sulfate (EMD), plus D-glucose (EMD), D-galactose (Sigma), or raffinose (Sigma). Cultures were grown in a humidified incubator

(Infors Multitron) at 30°C with rotary shaking at 230rpm (tubes and flasks) or 999rpm (600uL cultures in 1mL 96-well plates).

Flow cytometry assay in glucose gradient

GAL induction experiments were performed in a 2-fold dilution series of glucose concentration, from 2^{0}% to 2^{-9}% w/v, with constant 0.25% galactose. 2% glucose and 2% galactose conditions were also included with each glucose titration experiment. To assess and control for well-to-well variation, experiments were performed as a co-culture of a "query" strain to be phenotyped and a "reference" strain that was always SLYB93 (natural isolate YJM978 with constitutive mCherry segmentation marker).

To start an experiment, cells were struck onto YPD agar from -80C glycerol stocks, grown to colonies, and then inoculated from colony into YPD liquid and cultured for 16-24 hours. Then, query and reference strain cultures were mixed 9:1 by volume and inoculated in a dilution series (1:200 to 1:6400) in S + 2% raffinose medium. The raffinose outgrowths were incubated for 16-20 hours, and then their optical density (OD_{600}) was measured on a plate reader (PerkinElmer Envision). One outgrowth culture with OD_{600} closest to 0.1 was selected for each strain, and then washed once in S (with no carbon sources). Washed cells were diluted 1:200 into glucose + galactose gradients in 96-well plates (600uL cultures in each well) and incubated for 8 hours. Then, cells were harvested and fixed by washing twice in Tris-EDTA pH 8.0 (TE) and resuspended in TE + 0.1% sodium azide before transferring to a shallow microtiter plate (CELLTREAT) for measurement. Flow cytometry was performed using a Stratedigm S1000EX with A700 automated plate handling system. Data analysis was performed using custom MATLAB scripts, including Flow-Cytometry-Toolkit (https://github.com/springerlab/Flow-Cytometry-Toolkit).

Experiments using glycerol and acetate as pre-induction carbon sources were done as above, except S + 3% glycerol or S + 2% potassium acetate were used instead of raffinose medium for the outgrowth step.

Crossing and generating segregants

To prepare parent strains for crossing and sporulation, we sporulated diploid natural isolates bearing the hoΔ::GAL1pr-YFP-hphNT1 reporter cassette and isolated random spores that displayed MAT**a** or MATα phenotypes in test crosses. We then introduced a constitutive fluorescent marker in tandem with the GAL reporter, to obtain MAT**a**; hoΔ::GAL1pr-YFP-mTagBFP2-kanMX4 or MATα; hoΔ::GAL1pr-YFP-mCherry-natMX4 parent strains. To the MAT**a** parent we also introduced a pRS413-derived plasmid bearing STE2pr-AUR1-C and hphNT1. This plasmid is maintained by hygromycin selection but also allows selection for MAT**a** cells by Aureobasidin A [56]. This plasmid design is inspired by a similar mating-type selection plasmid used in a recent study [57].

To isolate segregants for phenotyping, we crossed a parent with BFP-kanMX + MAT-selection plasmid to a parent with mCherry-natMX and isolated a $G418^{R}Nat^{R}Hyg^{R}$ diploid hybrid with the plasmid. We sporulated the hybrid by culturing it to saturation in YPD, diluting 1:10 in YP+2% potassium acetate and incubating at 30C for 8 hours, and washing and resuspending into 2% potassium acetate and incubating at 30C until >20% of cells were tetrads, or about 3 days. We incubated ~$5x10^{6}$ tetrads in 100uL water with 50U of zymolyase 100T (Zymo Research) for 5 hours at 30C, and then resuspended tetrads in 1mL of 1.5% NP-40 and sonicated for 10 seconds at power setting 3 on a probe sonicator (Fisher Scientific). The resulting segregants were plated on YPD + 0.5ug/mL Aureobasidin A ("AbA"; Clontech) and random colonies were picked into

YPD liquid and saved as glycerol stocks. Haploidy was confirmed by mating to tester strains with known mating type. 90 segregants were phenotyped for GAL induction as described above.

Sorting-based bulk-segregant analysis

To generate segregant pools, we prepared a diploid hybrid and sporulated it as described above. To reduce the size of recombination blocks and improve the resolution of linkage mapping [58], we then performed the following "intercross" protocol 4 times: from spore suspension, use Sony SH800 Cell Sorter to sort $4x10^6$ BFP+ or mCherry+ (but not +/+ or -/-) cells into 100uL YPD + 40ug/mL tetracycline; incubate for 16 hours at 30C without shaking; add 5mL YPD + 200ug/mL G418 + 100ug/mL ClonNat + 200ug/mL Hygromycin B and incubate 48 hours at 30C with shaking; sporulate cultures and prepare sonicated spore suspension. After the 4^{th} sporulation cycle, the sonicated spores were resuspended in YPD + 0.5ug/mL AbA and incubated at 30C for 16 hours. This culture was frozen as a glycerol stock, as well as used to inoculate the galactose-induction sorting experiment.

To sort segregant pools for bulk genotyping, we inoculated the intercrossed, MAT**a**-selected segregants from a saturated YPD culture into S + 2% raffinose + AbA at dilutions of 1:200, 1:400, 1:800, and 1:1600, and incubated at 30C for 16-24 hours. We chose the raffinose culture with OD closest to 0.1, washed once in S (0.17% Yeast Nitrogen Base + 0.5% Ammonium Sulfate), and diluted 1:200 into S + 0.25% glucose + 0.25% galactose + AbA. We incubated the glucose-galactose culture at 30C for 8 hours, and then used a Sony SH800 sorter to isolate pools of 30,000 cells with the 5% lowest ("OFF") and highest ("HI") YFP expression, among cells whose Back Scatter (BSC) signal was between 10^5 and $3x10^5$. The "LOW" pool was similarly obtained, but from the 5% of cells with lowest non-basal expression (Figure S3). The sorted cells were resupended in YPD + AbA and incubated at 30C until saturation, about 16-24 hours. An

353 aliquot of this culture was saved for -80C glycerol stocks, and another was used to prepare
354 sequencing libraries.

355 To sequence the segregant pools, we extracted genomic DNA from 0.5mL of saturated YPD
356 culture of each segregant pool using the PureLink Pro 96 kit (Thermo Fisher K182104A). From
357 these genomic preps, we made sequencing libraries using Nextera reagents (Illumina FC-121-
358 1030) following a low-volume protocol [59]. We adjusted the input DNA concentration so that
359 resulting libraries had mean fragment sizes of 200-300bp as measured on a BioAnalyzer.
360 Libraries were multiplexed and sequenced in an Illumina NextSeq flow cell to a depth of 16-33x.

361 Reads from the Illumina sequencing were aligned to the sacCer3 reference genome using bwa
362 mem, and SNP counts were generated using samtools mpileup, on the Harvard Medical School
363 Orchestra cluster . These outputs were processed in MATLAB using custom code as follows:
364 SNPs with coverage less than 2 or more than 1000 were removed. The LOW and HI pools were
365 computationally merged into an ON pool. To make sure the two pools contributed equally to the
366 merged pool, at each SNP, allele counts in the pool with higher coverage were randomly
367 subsampled to the coverage of the other pool. The final allele counts in each pool were output to
368 text files by chromosome and given as inputs to the MULTIPOOL algorithm (mp_inference.py
369 version 0.10.2) [60] to compute LOD scores. Loci with maximal LOD>5 were considered
370 significant; previous work showed that this corresponded to an FDR of 5% [61,62]. This
371 correspondence may differ under our experimental conditions; therefore, the 2 loci that we did
372 not validate experimentally should be interpreted with caution.

373 **CRISPR/Cas9 allele replacement**

374 Allele replacement strains were constructed using 3 rounds of gene knockout followed by
375 CRISPR/Cas9-mediated markerless integration of heterologous allele. Initially, strains were
376 prepared by introducing Cas9 on a CEN/ARS plasmid (SLVF11); this plasmid is derived from a
377 previous one [63], but we replaced the auxotrophic *URA3* marker with AUR1-C to allow
378 Aureobasidin A selection on prototrophic natural isolates. In each round of allele replacement, a
379 gene plus upstream and downstream flanking sequences (-784bp to +815bp for *GAL3*, -449bp to
380 +372bp for *MKT1*, -191bp to +139bp for *GAL4*) was deleted by integration of a kanMX6 marker
381 with 40bp flanking homology. Then, a donor DNA, a guide RNA insert, and a guide RNA
382 backbone were simultaneously transformed into the strain [64]. The donor DNA contains the
383 new allele, its flanking sequences, and an additional 40bp of homology to target it to the correct
384 genomic locus. The guide RNA insert was a linear DNA containing a SNR52 promoter driving a
385 guide RNA gene containing a 20bp CRISPR/Cas recognition sequence linked to a crRNA
386 scaffold sequence, plus 40bp of flanking homology on both ends to a guide RNA backbone. The
387 guide RNA backbone was a 2u plasmid containing natMX4 (pRS420). This was linearized by
388 NotI + XhoI digestion before transformation. Allele re-integration transformations were plated
389 on cloNAT to select for in vivo assembly of the guide RNA into a maintainable plasmid, and
390 Aureobasidin A to select for presence of Cas9. Successful re-integration was verified by colony
391 PCR and Sanger sequencing was performed on a subset of strains and on all donor DNAs to
392 verify the sequence of allelic variants.

Acknowledgments

Rishi Jajoo, Dawn Thompson, and Andrew Murray for critical reading of the manuscript; Sarah Boswell, Bo Hua, Yarden Katz, and Chiara Ricci-Tam for helpful discussions; Shervin Javadi and Stratedigm for flow cytometry assistance.

References

1. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005;6: 451–464. doi:10.1038/nrg1615

2. Balázsi G, van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: from microbes to mammals. Cell. 2011;144: 910–25. doi:10.1016/j.cell.2011.01.030

3. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell. 2008;135: 216–226. doi:10.1016/j.cell.2008.09.050

4. Grimbergen AJ, Siebring J, Solopova A, Kuipers OP. Microbial bet-hedging: the power of being different. Curr Opin Microbiol. 2015;25: 67–72.

5. Veening J-W, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol. 2008;62: 193–210. doi:10.1146/annurev.micro.62.081307.163002

6. Xiong W, Ferrell JEJ. A positive-feedback-based bistable “memory module” that governs a cell fate decision. Nature. 2003;426: 460–465. doi:10.1038/nature02119.1.

7. Macarthur BD, Ma’ayan A, Lemischka IR. Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Biol. Nature Publishing Group; 2009;10: 672–681. doi:10.1038/nrm2766

8. Johnston M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev. 1987;51: 458.

9. Bhat PJ. Galactose Regulon of Yeast: From Genetics to Systems Biology. Berlin: Springer; 2008.

10. Lohr D, Venkov P, Zlatanova J, Program B, Academy B. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. 1995;9: 777–787.

11. Escalante-chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ, et al. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proc Natl Acad Sci U S A. 2015;112: 1636–1641. doi:10.1073/pnas.1418058112

12. Biggar SR, Crabtree GR. Cell signaling can direct either binary or graded transcriptional

responses. EMBO J. 2001;20: 3167–76. doi:10.1093/emboj/20.12.3167

13. Venturelli OS, El-Samad H, Murray RM. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. Proc Natl Acad Sci U S A. 2012;109: 1–10. doi:10.1073/pnas.1211902109

14. Acar M, Becskei A, Oudenaarden A van. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228. doi:10.1038/nature03524

15. Hawkins KM, Smolke CD. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. J Biol Chem. 2006;281: 13485–92. doi:10.1074/jbc.M512317200

16. Acar M, Pando BF, Arnold FH, Elowitz MB, van Oudenaarden A. A general mechanism for network-dosage compensation in gene circuits. Science (80-). 2010;329: 1656–60. doi:10.1126/science.1190544

17. Ramsey S a, Smith JJ, Orrell D, Marelli M, Petersen TW, de Atauri P, et al. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. Nat Genet. 2006;38: 1082–7. doi:10.1038/ng1869

18. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat Commun. 2015;6: 6226.

19. Lee KB, Wang J, Escalante-Chong R, Palme J, Springer M. Natural variation in S. cerevisiae glucose-galactose sensing is driven by polymorphisms in the signal transducer Gal3p. Unpublished. 2016;

20. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. PLoS Biol. 2015;13: e1002042. doi:http://dx.doi.org/10.1101/002907

21. Wang J, Atolia E, Hua B, Savir Y, Escalante-chong R. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost – Benefit Tradeoff. PLoS Biol. 2015; 1–31. doi:10.1371/journal.pbio.1002041

22. New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, et al. Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments. Doebeli M, editor. PLoS Biol. 2014;12: e1001764. doi:10.1371/journal.pbio.1001764

23. Solopova A, van Gestel J, Weissing FJ, Bachmann H, Teusink B, Kok J, et al. Bet-hedging during bacterial diauxic shift. Proc Natl Acad Sci U S A. 2014;111: 7427–32. doi:10.1073/pnas.1320063111

24. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James S a, et al. Population genomics of domestic and wild yeasts. Nature. Nature Publishing Group; 2009;458: 337–41. doi:10.1038/nature07743

25. Cromie G a, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P, et al.

Genomic sequence diversity and population structure of Saccharomyces cerevisiae assessed by RAD-seq. G3 (Bethesda). 2013;3: 2163–71. doi:10.1534/g3.113.007492

26. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics. BioMed Central Ltd; 2012;13 Suppl 6: S8.

27. Gancedo JM. Yeast Carbon Catabolite Repression. Microbiol Mol Biol Rev. 1998;62.

28. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. Science. 2000;290: 2306–9. doi:10.1126/science.290.5500.2306

29. Hidalgo P, Ansari AZ, Schmidt P, Hare B, Simkovich N, Farrell S, et al. Recruitment of the transcriptional machinery through GAL11P: Structure and interactions of the GAL4 dimerization domain. Genes Dev. 2001;15: 1007–1020. doi:10.1101/gad.873901

30. Hong M, Fitzgerald MX, Harper S, Luo C, Speicher DW, Marmorstein R. Structural Basis for Dimerization in DNA Recognition by Gal4. Structure. 2008;16: 1019–1026. doi:10.1016/j.str.2008.03.015

31. Himmelfarb HJ, Pearlberg J, Last DH, Ptashne M. GAL11P□: A Yeast Mutation That Potentiates the Effect of Weak GAL4-Derived Activators. Cell. 1990;63: 1299–1309.

32. Wickner RB. Plasmids Controlling Exclusion of the K2 Killer Double-Stranded RNA Plasmid of Yeast. Cell. 1980;21.

33. Tadauchi T, Inada T, Matsumoto K, Irie K. Posttranscriptional regulation of HO expression by the Mkt1-Pbp1 complex. Mol Cell Biol. 2004;24: 3670–3681.

34. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro J a, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. Nature Publishing Group; 2010;464: 1039–1042.

35. Deutschbauer AM, Davis RW. Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet. 2005;37: 1333–40. doi:10.1038/ng1674

36. Demogines A, Smith E, Kruglyak L, Alani E. Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. PLoS Genet. 2008;4: e1000123.

37. Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE. Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of Saccharomyces cerevisiae {S288C} strains. Genetics. 2009;183: 365–383.

38. Kim HS, Fay JC. A combined-cross analysis reveals genes with drug-specific and background-dependent effects on drug sensitivity in Saccharomyces cerevisiae. Genetics. 2009;183: 1141–1151.

39. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, et al. Dissecting the architecture of a quantitative trait locus in yeast. Nature. 2002;416: 326–

330.

40. Lee S-I, Dudley AM, Drubin D, Silver PA, Krogan NJ, Pe'er D, et al. Learning a prior on regulatory potential from eQTL data. PLoS Genet. 2009;5: e1000358.

41. Gerber AP, Herschlag D, Brown PO. Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. Sean Eddy, editor. PLoS Biol. Public Library of Science; 2004;2: e79. doi:10.1371/journal.pbio.0020079

42. Hua B, Springer M. Hierarchical structure explains why a quarter of the yeast genome affects a quantitative trait.

43. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. 2013;494: 234–7. doi:10.1038/nature11867

44. Hou J, Sigwalt A, Fournier T, Pflieger D, Peter J, de Montigny J, et al. The Hidden Complexity of Mendelian Traits across Natural Yeast Populations. Cell Rep. 2016;

45. Johnston M, Flick JS, Pextont T. Multiple Mechanisms Provide Rapid and Stringent Glucose Repression of GAL Gene Expression in Saccharomyces cerevisiae. 1994;14. doi:10.1128/MCB.14.6.3834.Updated

46. Carlson M, Osmond BC, Botstein D. Mutants of yeast defective in sucrose utilization. Genetics. 1981;98.

47. Zimmermann, Friedrich K Entian K-D. Yeast sugar metabolism. CRC Press; 1997.

48. Huberts DHEW, Niebel B, Heinemann M. A flux-sensing mechanism could regulate the switch between respiration and fermentation. FEMS Yeast Res. 2012;12: 118–128. doi:10.1111/j.1567-1364.2011.00767.x

49. Turcotte B, Liang XB, Robert F, Soontorngun N. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Res. 2010;10: 2–13. doi:10.1111/j.1567-1364.2009.00555.x

50. DeRisi JL. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. Science (80-). 1997;278: 680–686. doi:10.1126/science.278.5338.680

51. van Dijken JP, Scheffers WA. Redox balances in the metabolism of sugars by yeasts. FEMS Microbiol Lett. 1986;32: 199–224. doi:10.1016/0378-1097(86)90291-0

52. Stockwell SR, Landry CR, Rifkin S a. The yeast galactose network as a quantitative model for cellular memory. Mol Biosyst. Royal Society of Chemistry; 2014;

53. Kumar PR, Yu Y, Sternglanz R, Johnston SA, Joshua-Tor L, Rajesh Kumar P, et al. NADP Regulates the Yeast GAL Induction System. Science (80-). 2008;319: 1090–2. doi:10.1126/science.1151903

54. Li Y, Chen G, Liu W. Multiple metabolic signals influence GAL gene activation by

modulating the interaction of Gal80p with the transcriptional activator Gal4p. Mol Microbiol. 2010;78: 414–428. doi:10.1111/j.1365-2958.2010.07343.x

55. Healey D, Axelrod K, Gore J. Negative frequency-dependent interactions can underlie phenotypic heterogeneity in a clonal microbial population. Mol Syst Biol. 2016;12: 877.

56. Hashida-Okado T, Ogawa A, Kato I, Takesako K. Transformation system for prototrophic industrial yeasts using the AUR1 gene as a dominant selection marker. FEBS Lett. 1998;425: 117–122. doi:10.1016/S0014-5793(98)00211-7

57. Treusch S, Albert FW, Bloom JS, Kotenko IE, Kruglyak L. Genetic Mapping of MAPK-Mediated Complex Traits Across S. cerevisiae. PLoS Genet. 2015;11: e1004913. doi:10.1371/journal.pgen.1004913

58. Cubillos F a, Parts L, Salinas F, Bergström A, Scovacricchi E, Zia A, et al. High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. Genetics. 2013;195: 1141–55. doi:10.1534/genetics.113.155515

59. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLoS One. 2015;10: e0128036. doi:10.1371/journal.pone.0128036

60. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics. BioMed Central Ltd; 2012;13 Suppl 6: S8. doi:10.1186/1471-2105-13-S6-S8

61. Treusch S, Albert FW, Bloom JS, Kotenko IE, Kruglyak L. Genetic Mapping of {MAPK-Mediated} Complex Traits Across S. cerevisiae. PLoS Genet. 2015;11: e1004913.

62. Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. Genetics of single-cell protein abundance variation in large yeast populations. Nature. Nature Publishing Group; 2014;506: 494–497. doi:10.1038/nature12904

63. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. 2013;41.

64. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, et al. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Syst. Elsevier; 2015; 1–9. doi:10.1016/j.cels.2015.02.001

Figure Captions

Figure 1. Natural variation in GAL induction can be analyzed in terms of two uncorrelated features

Each plot is a series of YFP fluorescence (normalized to side scatter "SSC") histograms from 12 sugar conditions for strains (A) I14, (B) S288C, (C) DBVPG1106, and (D) YJM978. Other phenotyped strains are shown in Figure S1. Darker regions represent more frequently observed YFP values. The middle 10 conditions in each plot are 0.25% galactose + the indicated concentrations of glucose. The first and last conditions contain only one sugar: "D", 2% glucose; "G", 2% galactose. (E) Identification of induced cell subpopulation (green shading) using a reference distribution from 2% glucose (black histogram) (Materials and Methods). (F) Induced level (blue line) and induced fraction (orange line), and the corresponding E_{10} and F_{50} metrics, for strain DBVPG1106. (G) Scatterplot of E_{10} versus F_{50} across 34 *S. cerevisiae* natural isolates (mean and S.D.; n=3-10).

Figure 2. Bulk segregant analysis of E_{10} and F_{50}

(A) E_{10} versus F_{50} across 90 haploid segregants of the DBVPG1106 x S288C cross. Parent phenotypes are shown as filled circles: DBVPG1106 (red), S288C (blue). (B) Schematic of bulk segregant analysis strategy. (C) GAL reporter histograms of parent strains DBVPG1106 (red) and S288C (blue) and a pool of haploid segregants (gray, bottom) in the sorting conditions, 0.25% glucose + 0.25% galactose. Green boxes are a schematic of the gates used to sort segregant cells into 3 phenotyped pools for sequencing (Gates used in actual sorting experiment are shown in Figure S3). ON pool allele counts are a computational sum of the LOW and HI pool allele counts (Materials and methods). (D) Genome-wide plots of differential allele frequency and log-odds-ratio (LOD) as computed by the MULTIPOOL algorithm (see Main Text, Materials and Methods). Top two panels show the OFF/ON comparison; bottom two panels show the LOW/HIGH comparison.

Figure 3. Combinatorial effects of strain background and *GAL3*, *MKT1*, and *GAL4* alleles.

(A) E_{10} versus F_{50} for all 16 combinations of S288C ("S") or DBVPG1106 ("D") strain background (gray letters), *GAL3* allele (red), *MKT1* allele (green), and *GAL4* allele (blue). Effects of switching from DBVPG1106 to S288C variant while holding other genetic variables

589 constant are shown as arrows for switching (B) *GAL3* allele, (C) *MKT1* allele, (D) *GAL4* allele,
590 or (E) strain background. (F) The effects shown in (B)-(E) are plotted as differences in E_{10}
591 versus differences in F_{50}.

592 **Figure 4. Perturbations that affect F_{50} do not affect E_{10}**
593 (A) F_{50} and E_{10} for allele-replacement strains with S288C, BC187, or DBVPG1106 genetic
594 backgrounds but containing alleles of GAL3 from various other natural isolates. (B) F_{50} and E_{10}
595 for 8 natural isolate strains induced in glucose+galactose after being cultured in raffinose,
596 glycerol, or acetate. Raffinose pre-culture is the standard condition used for the other
597 experiments in this paper.

598 **Figure 5. Changing F_{50} changes the number of modes of the GAL response**
599 Plotted are GAL reporter histogram series on a glucose gradient + galactose, as in Figure 1F,
600 with induced expression level (blue line), E_{10} (dotted vertical blue line), induced fraction (orange
601 line), and F_{50} (dotted vertical orange line). These plots show one representative experiment for
602 each strain/condition, out of the 3-12 replicates plotted in Figures 4.3 and 4.4. Strains: (A)
603 DBVPG1106 with all endogenous alleles; (B) DBVPG1106 with replacements by *GAL3* and
604 *MKT1* alleles from S288C. (C) BC187 with endogenous alleles; (D) BC187 with replacement by
605 *GAL3* allele from YJM978; (E) Y12-WashU cultured in raffinose prior to glucose + galactose
606 (standard protocol); (F) Y12-WashU pre-cultured in acetate.

Supporting Information

Text S1. All Supporting Figures

Contains Figures S1-4 and their captions.

Figure S1. GAL response phenotypes for 34 natural isolates

Plotted are series of GAL1pr-YFP fluorescence (normalized to side scatter "SSC") histograms from 12 sugar conditions for 34 strains. One replicate experiment (out of 3-10 replicates) is shown for each strain. Data from all replicates is used to calculate E_{10} and F_{50} for the scatterplot in Figure 1G.

Figure S2. Bimodal phenotypes simulated using a subpopulation decomposition framework

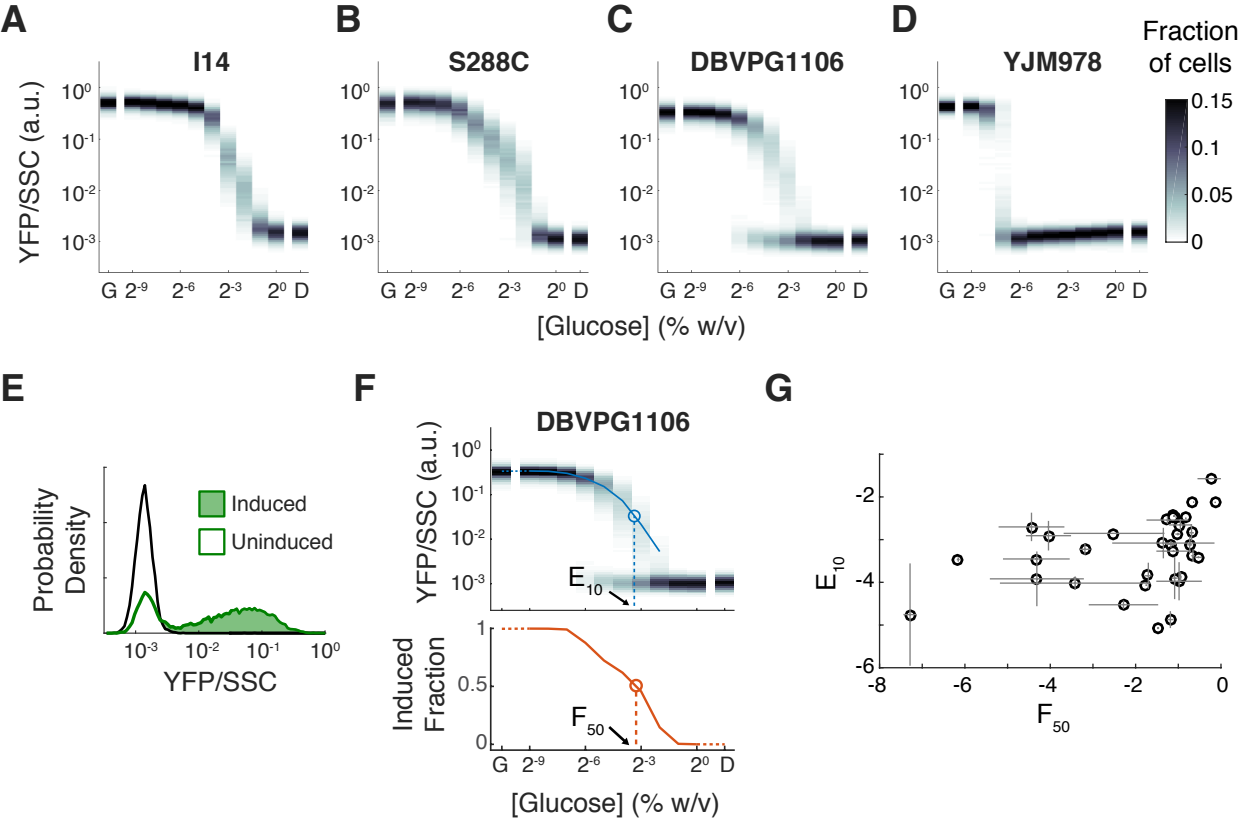
(A) Two simulated subpopulations, where the mean of the induced population is shown in blue. (B) 3 possible functions for the dependence of induced fraction on glucose. (C) Simulated population behaviors using the 3 induced fraction functions.

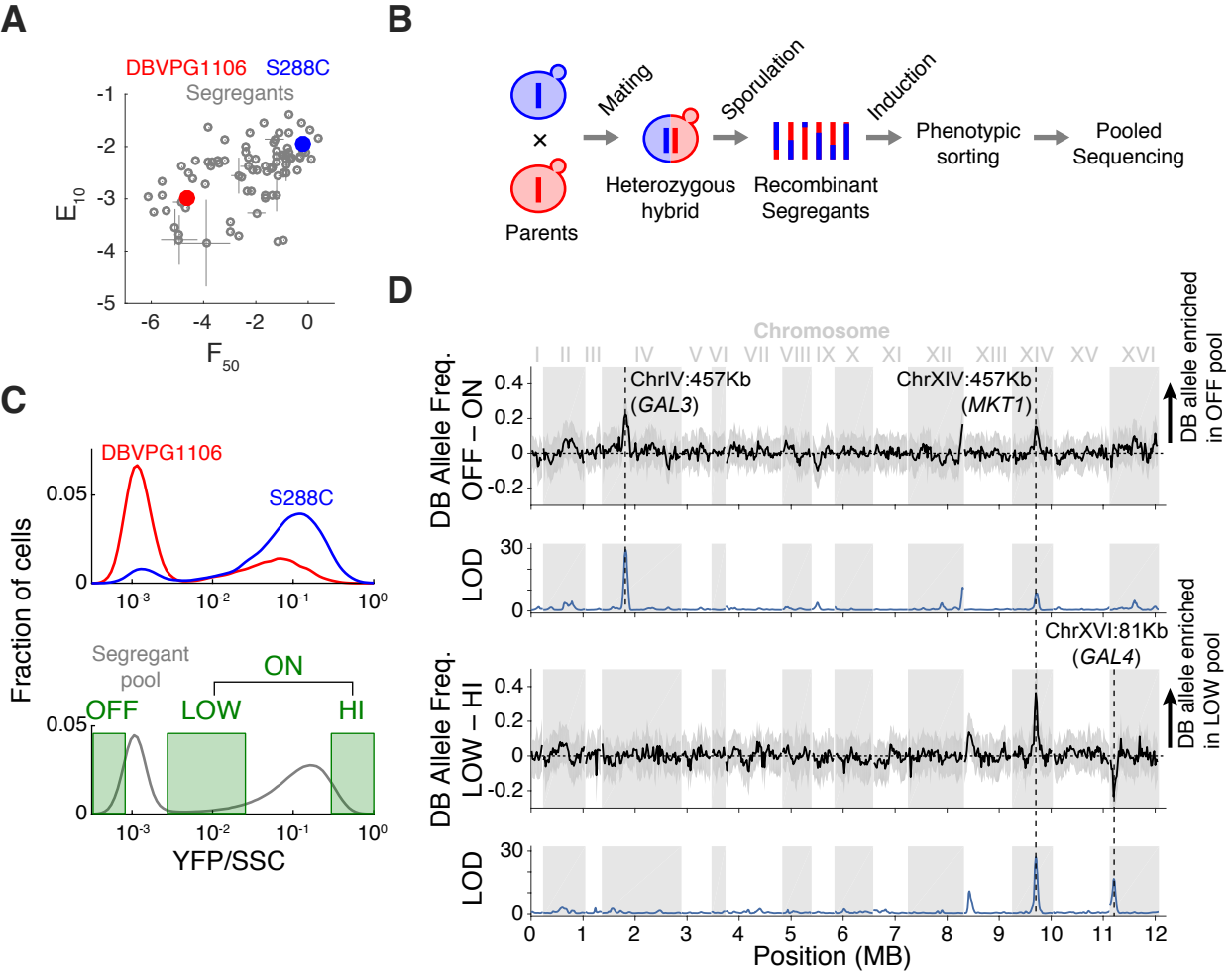
Figure S3. Sorting strategy for bulk-segregant analysis

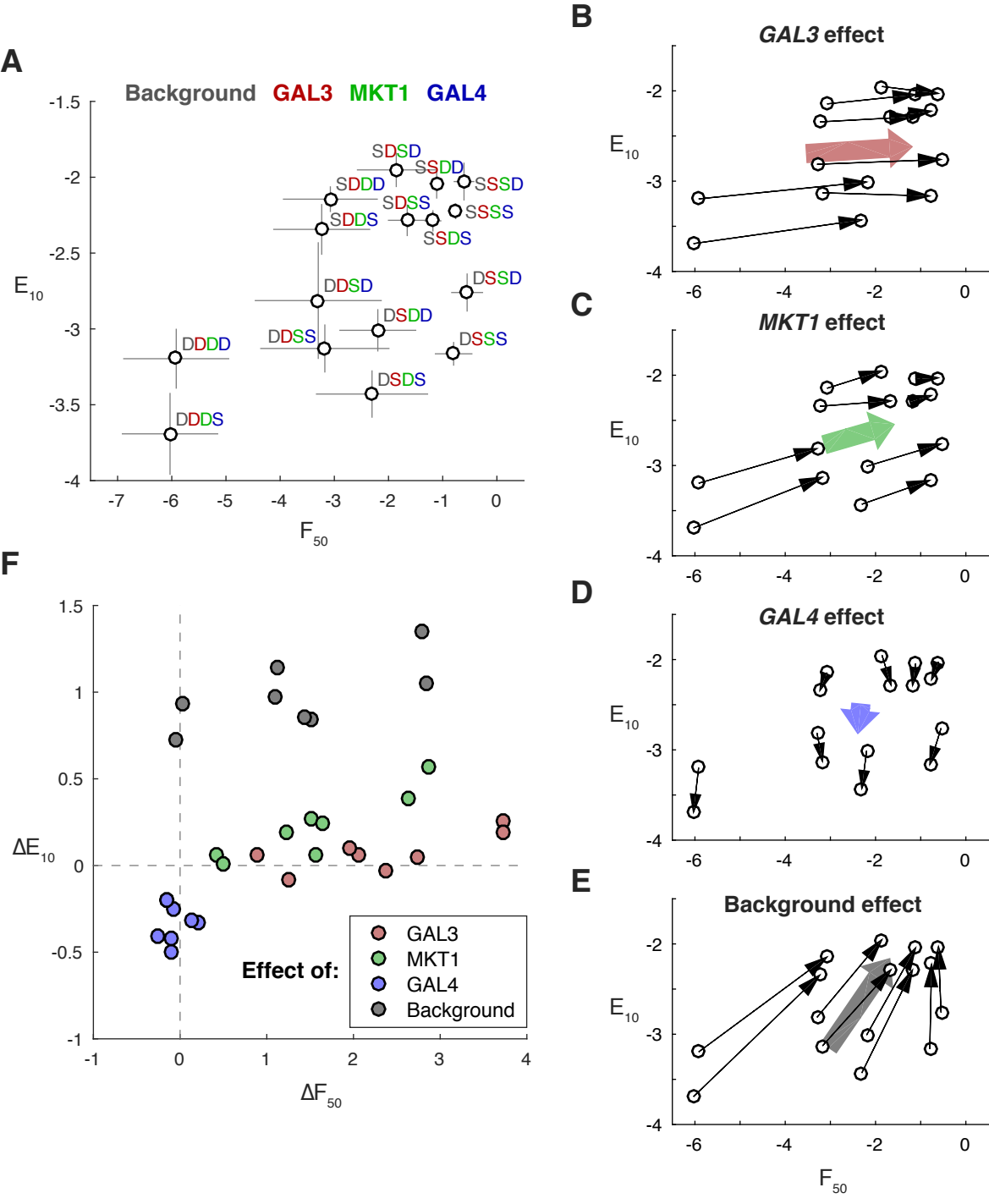
(A) Backscatter versus forward scatter of unsorted segregant pool, obtained on Sony SH800 cell sorter. (B) Backscatter versus FITC (YFP), showing mixture of uninduced and induced cells. Backscatter was used as a proxy for cell size; therefore, it is correlated with fluorescence. Gating on backscatter (rectangle) isolates differences in GAL1pr-YFP reporter among the cells. (C) Gates for OFF, LOW, and HI cells were drawn after gating on backscatter and shaped to follow the backscatter-FITC correlation. (D) View of gated populations as histogram on FITC axis.

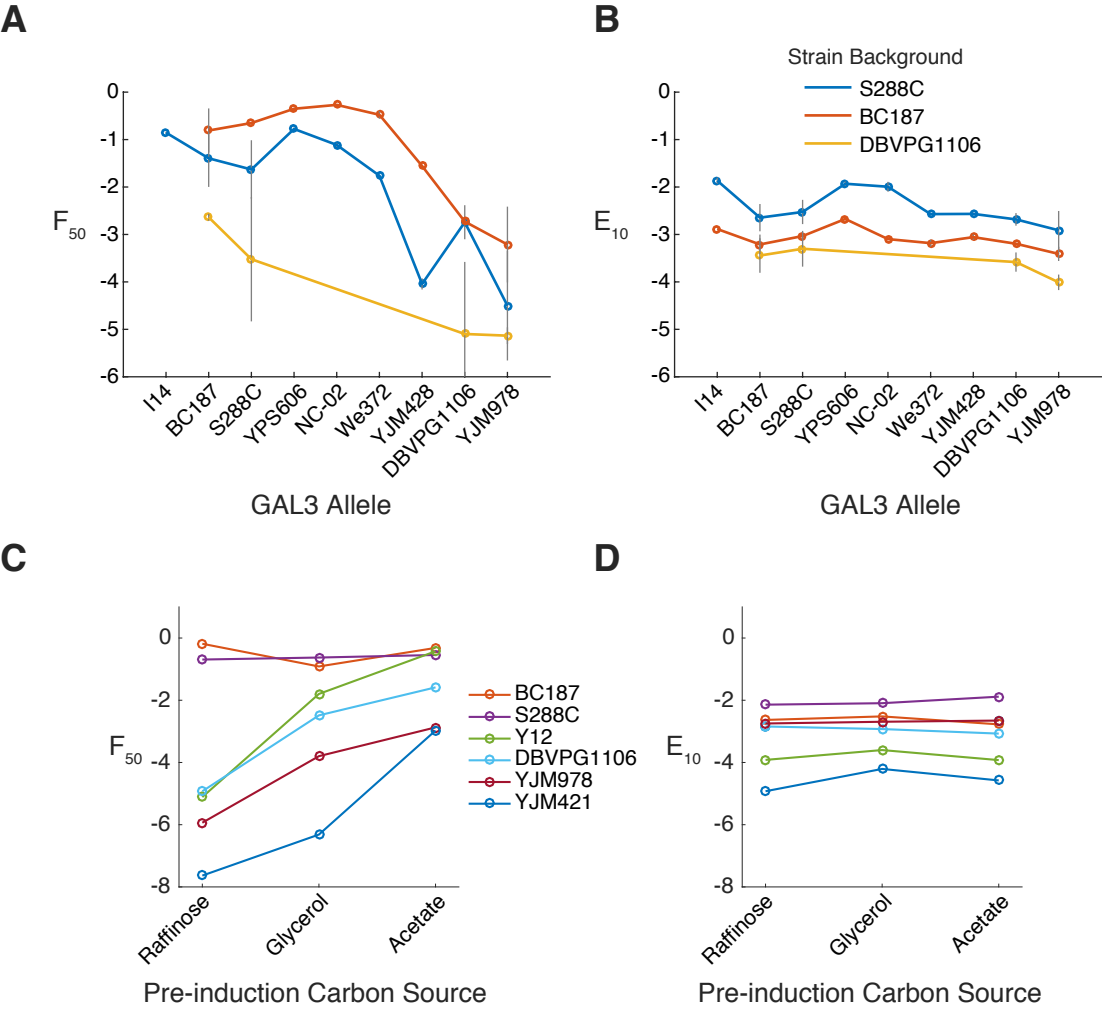
Figure S4. Effect of allele replacement of *GAL3*, *MKT1*, or *GAL4* in DBVPG1106 and S288C backgrounds.

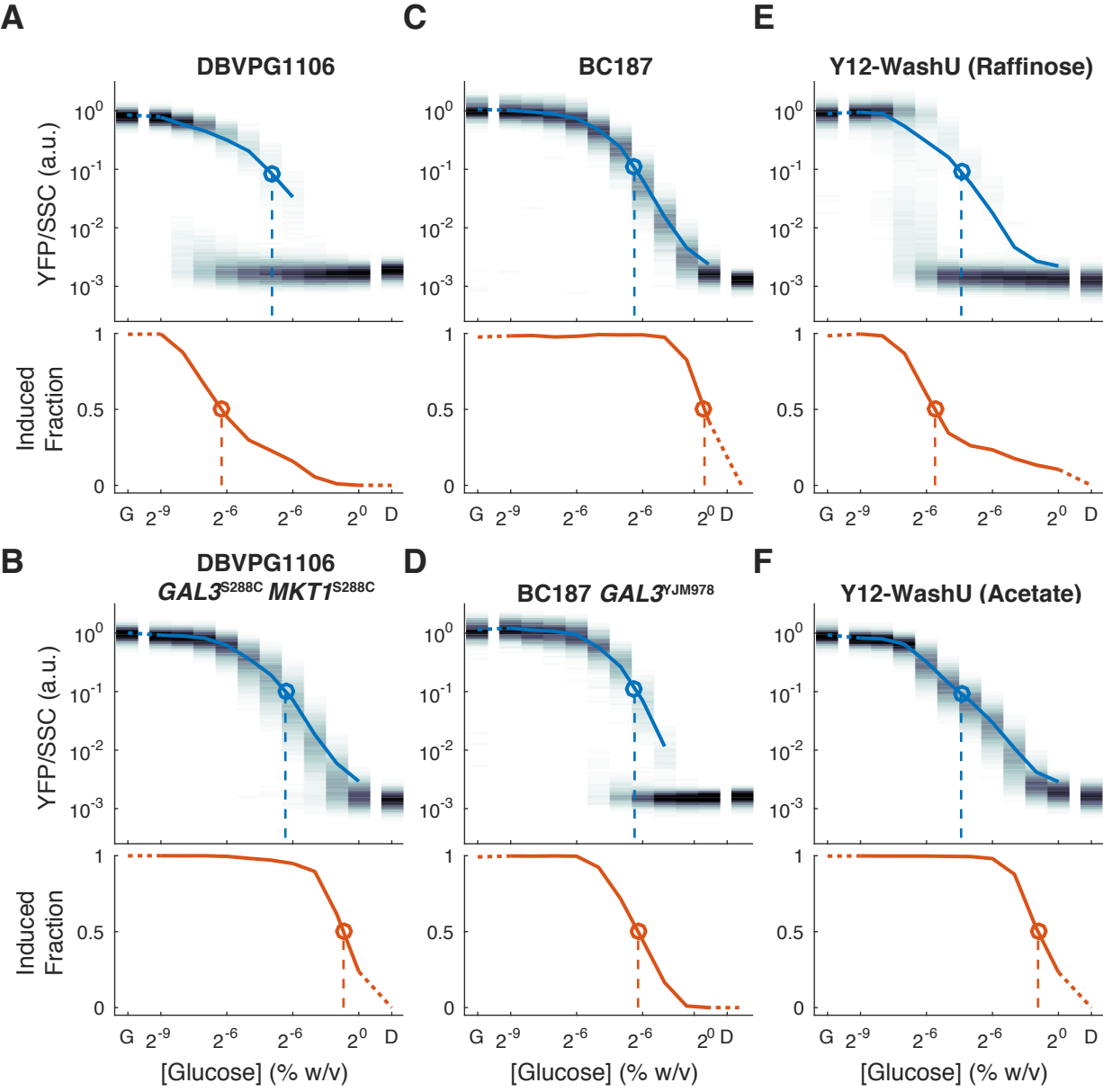
Scatterplots of E_{10} versus F_{50} for (A) DBVPG1106 strains where the indicated genes (and flanking regions) have been replaced by their S288C alleles; (B) S288C strains containing replacements by DBVPG1106 alleles. (C) Enlargement of region in (B) outlined by dotted rectangle. Small circles are individual replicates (12 replicates per genotype, comprising 6 replicates each for 2 independently constructed isolates – see Materials and Methods); large circles indicate the mean. These plots show a subset of the same data as in Figure 3.


A
I14
B
S288C
C
DBVPG1106
D
YJM978
Fraction of cells
0.15
0.1
0.05
0
YFP/SSC (a.u.)
10^0
10^{-1}
10^{-2}
10^{-3}
G 2^{-9} 2^{-6} 2^{-3} 2^0 D
[Glucose] (% w/v)
E
Probability Density
Induced
Uninduced
10^{-3} 10^{-2} 10^{-1} 10^0
YFP/SSC
F
DBVPG1106
YFP/SSC (a.u.)
E_{10}
Induced Fraction
1
0.5
0
F_{50}
G 2^{-9} 2^{-6} 2^{-3} 2^0 D
[Glucose] (% w/v)
G
E_{10}
-2
-4
-6
-8 -6 -4 -2 0
F_{50}


A
DBVPG1106
S288C
Segregants
E_{10}
F_{50}
B
Mating
Sporulation
Induction
Phenotypic sorting
Pooled Sequencing
Parents
Heterozygous hybrid
Recombinant Segregants
C
DBVPG1106
S288C
Fraction of cells
Segregant pool
OFF
ON
LOW
HI
YFP/SSC
D
Chromosome
DB Allele Freq. OFF – ON
ChrIV:457Kb (GAL3)
ChrXIV:457Kb (MKT1)
DB allele enriched in OFF pool
LOD
DB Allele Freq. LOW – HI
ChrXVI:81Kb (GAL4)
DB allele enriched in LOW pool
Position (MB)


A
Background GAL3 MKT1 GAL4
SDSD
SSDD
SDDD
SSSD
SDSS
SSSS
SDDS
SSDS
DDSD
DSSD
DSDD
DDSS
DSSS
DDDD
DSDS
DDDS
E10
F50
B
GAL3 effect
C
MKT1 effect
D
GAL4 effect
E
Background effect
F
ΔE10
ΔF50
Effect of:
GAL3
MKT1
GAL4
Background


A
B
C
D
F_{50}
E_{10}
Strain Background
S288C
BC187
DBVPG1106
I14
BC187
S288C
YPS606
NC-02
We372
YJM428
DBVPG1106
YJM978
GAL3 Allele
BC187
S288C
Y12
DBVPG1106
YJM978
YJM421
Raffinose
Glycerol
Acetate
Pre-induction Carbon Source


A
DBVPG1106
B
DBVPG1106
GAL3S288C MKT1S288C
C
BC187
D
BC187 GAL3YJM978
E
Y12-WashU (Raffinose)
F
Y12-WashU (Acetate)
YFP/SSC (a.u.)
Induced Fraction
G
D
[Glucose] (% w/v)