The shape of fitness functions and the distribution of mutational effect sizes jointly limit adaptation by regulatory mutations

Simon Aubé[1,2,3,4,5*], Alexandre K. Dubé[1,2,3,4,5] & Christian R. Landry[1,2,3,4,5*]

1. Département de Biochimie, microbiologie et bio-informatique, Université Laval, 1045 Avenue de la Médecine, Québec, QC, Canada, G1V 0A6
2. Institut de Biologie Intégrative et des Systèmes (IBIS), 1030, Avenue de la Médecine, Université Laval, Québec, QC, Canada, G1V 0A6
3. PROTEO-Regroupement Québécois de Recherche sur la Fonction, l'Ingénierie et les Applications des Protéines, Université Laval, 1030, Avenue de la Médecine Québec, QC, Canada, G1V 0A6
4. Centre de Recherche en Données Massives de l'Université Laval, Université Laval, 1065, Avenue de la Médecine, Québec, QC, Canada, G1V 0A6
5. Centre de recherche en infectiologie, Université Laval, 2705 boul. Laurier, Québec, QC, G1V 4G2

Correspondence: simon.aube.2@ulaval.ca and christian.landry@bio.ulaval.ca

Abstract

Mutations in gene regulatory regions have been shown to play a role in rapid adaptation but the factors determining their contribution are largely unknown. Using the yeast metabolic enzyme cytosine deaminase, we examine if adaptation to 5-fluorocytosine (5-FC), which requires reduced cytosine deamination and can readily arise from amino acid substitutions, may be reached by single promoter mutations. We generated all single-nucleotide substitutions and indels in the *FCY1* promoter and assayed the resulting mutants in presence of 5-FC. This revealed that no promoter mutation is sufficient for adaptation to occur. We next investigated how this inaccessibility of adaptation arises by combining large-scale expression measurements with the experimental characterization of the corresponding expression-fitness function. These experiments showed that the shape of the corresponding fitness function precludes single promoter mutations from being adaptive. Although 24% of mutations significantly affect expression, the fitness curve is flat around wild-type level. As such, adaptation can only emerge from a severe reduction of expression, which cannot occur from a single mutation in the promoter. Our results show that the contribution of regulatory mutations to rapid adaptation not only depends on the distribution of mutational effect sizes on expression level but also on the shape of the function linking fitness to expression levels.

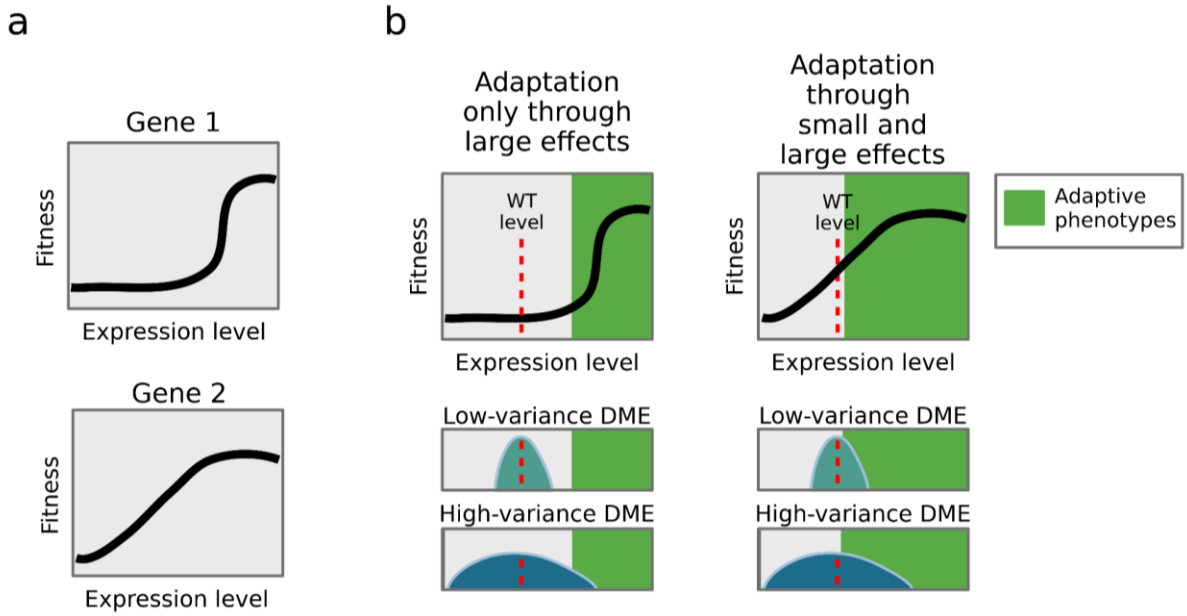
Introduction

For rapid adaptation to occur, advantageous phenotypes must be immediately available through mutation. Whether a given mutation is adaptive depends on how the corresponding molecular change translates into a fitness effect. For expression changes, this is defined by the affected gene's fitness function (**Fig. 1a**) – describing the relationship linking expression level to organismal fitness –, of which a diversity of shapes have been reported[1]. As such, the corresponding distribution of mutational effects (DME) on expression level and the shape of the relevant fitness function jointly dictate if adaptive changes can arise within a given cis-regulatory region. These two factors thus have a major impact on the available pathways to adaptation, as well as on the rate and mode of regulatory evolution. In spite of this importance, the interplay between them remains unknown for most genes[2]. Adaptation would for instance be unlikely to arise from the combination of a narrow DME, such that mutations have small effects on expression, with a step-like fitness function where only large expression changes are beneficial. Conversely, single-nucleotide mutations sampled from the same narrow DME could be adaptive if the fitness function instead displayed more gradual variation, allowing small expression changes to provide a benefit (**Fig. 1b**). Fully understanding the evolution of cis-regulatory sequences and its contribution to adaptation hence requires assaying mutational effects on expression level and characterizing their conversion through fitness functions[2].

Across organisms, a major engine of rapid adaptation is loss-of-function (LOF) mutations, both in coding and regulatory sequences. In humans, the best known mechanisms of resistance to malaria are the loss of expression of specific genes in red blood cells[3, 4] and the partial loss of a key metabolic enzyme activity[5]. Among the best documented adaptive morphological changes in vertebrates in natural populations arose through the repeated loss of expression of a developmental gene[6]. LOF mutations in coding sequences are also a major pathway to antimicrobial resistance, for instance in fungal pathogens evolving in response to antifungal drugs[7]. Expression-decreasing epigenetic changes are involved in the adaptive evolution of cancer cells, by reducing expression of tumor suppressor genes[8]. Finally, laboratory microbial evolution has shown that adaptation occurs through LOF at least as frequently as through gain of function(s)[9–16]. LOF therefore provides a powerful model to investigate questions regarding the contributions of mutations to rapid adaptation.

Adaptive LOF sometimes occurs in regulatory elements of genes and sometimes in coding sequences. While this relative contribution could be caused by chance only, some gene properties could make one type of change more likely to occur than another. However, past experiments are insufficient to answer this question. Large-scale studies which investigated the effect of mutations within promoters and cis-regulatory sequences[17–23] mostly focused on expression level, without directly measuring fitness effects. In few cases where the fitness consequences of promoter mutations are well known, for instance in *TDH3*[24–26] and *SUL1*[27] in yeast, only a

93 handful of coding sequence mutations have comparatively been assayed for the same
94 genes[28,29]. The contribution of regulatory mutations to rapid adaptation, relative to that
95 of coding changes, thus remains unclear.
96



97
98 **Fig. 1: The DME and the shape of the fitness function jointly define the**
99 **accessibility of adaptation. a,** Fitness functions describe the relationship between
100 the expression level of a gene and organismal fitness. Hypothetical fitness functions
101 are shown for two genes. In each case, the shape of the curve displays how fitness
102 varies with increasing expression level of the corresponding gene. **b,** The accessibility
103 of adaptation through expression level changes depends on mutational effect sizes
104 and fitness function shape. For the two fitness functions, adaptation from the same
105 low-fitness wild-type (WT) expression level is shown. When the function is flat around
106 WT expression (left), only large expression increases from the high-variance DME can
107 be adaptive. In contrast, when the fitness function is curved near WT expression (right),
108 small expression increases are adaptive, such that adaptation is accessible from the
109 low-variance DME as well as the high-variance one. Created in BioRender. Aubé, S.
110 (2025) https://BioRender.com/rw8qljy.
111
112 We focused on the adaptive loss of Fcy1 activity in yeast exposed to prodrug 5-
113 fluorocytosine (5-FC). While this enzyme's canonical activity is cytosine deamination,
114 it also converts 5-FC into 5-fluorouracil, which leads to growth arrest and cell death[30].
115 Accordingly, complete or partial LOF mutations in the *FCY1* coding sequence confer
116 resistance and are adaptive[31]. To assess whether regulatory mutations can also
117 contribute to rapid adaptation to this condition, we performed saturation mutagenesis
118 of the *FCY1* promoter. We measured the effect of all possible single nucleotide
119 substitutions and indels in this region on fitness and expression level. These
120 experiments revealed no measurable fitness gain in medium with 5-FC, although a
121 quarter of single-nucleotide changes significantly affect gene expression level.
122 Through the characterization of the underlying fitness function, we show that this
123 discrepancy emerges jointly from the molecular effect of mutations and from the shape

of this relationship. Because the fitness function is flat around the endogenous expression level, adaptation can only occur through large expression decreases, which cannot arise from a single-nucleotide change in the promoter. These findings highlight how the availability of adaptative phenotypes through mutation can be restricted by the combined action of the DME and of the relevant fitness function. In *FCY1*, this results in rapid adaptation being only accessible through mutations in the coding sequence. This observation may be generalizable to other enzymes which exert low control over metabolic fluxes.

Results

Measuring the effects of mutations within the *FCY1* promoter

To identify promoter regions to mutate, we performed deletions of successive 77 bp fragments of the intergenic region upstream of *FCY1*. This revealed that only the most proximal fragment, among four, is essential for gene function **(Extended Data Fig. 1**). For this reason, and because most regulatory elements in yeast promoters are found within the 150 bp closest to the transcribed gene[32,33], we focused on the 154 bp immediately upstream of *FCY1*. This corresponds to two fragments hereafter named F3 and F4 (**Fig. 2a**). Within this region, we generated all single-nucleotide mutations – substitutions, insertions and deletions. The corresponding variants of F3 and F4 were inserted at the endogenous *FCY1* locus using a CRISPR-Cas9-based pooled approach (**Fig. 2b**, *Methods*[34,35]). This resulted in 1021 unique mutated sequences (462 substitutions, 463 insertions and 96 deletions), of which 1020 were successfully recovered by sequencing of the mutant pools (**Supplementary Fig. 2**). This process was repeated four times for each of the three classes of mutations (**Fig. 2c**). The corresponding pools of substitutions, insertions and deletions were then combined, resulting in four independent transformation replicates, each containing all mutations. Because the set of substitutions was obtained by successively degenerating each nucleotide position, the WT *FCY1* promoter was also (re)constructed along with the desired mutants. Within each combined transformation replicate, its initial frequency was ~17%.

From each of the four transformation replicates, two culture replicates were made, which were both grown in liquid media with [12 μM] 5-FC and without, resulting in 16 cultures. This concentration has been shown to inhibit the growth of *Saccharomyces cerevisiae* by ~50% and was previously used to assay all single-codon substitutions in the *FCY1* coding sequence[31] (**Extended Data Fig. 2**). Additional replicates of all 16 cultures were also done by combining the promoter mutants with 960 representative coding sequence variants[31], to ensure that the fitness effects of promoter mutations were directly comparable to those of coding mutations(**Extended Data Fig. 3**).

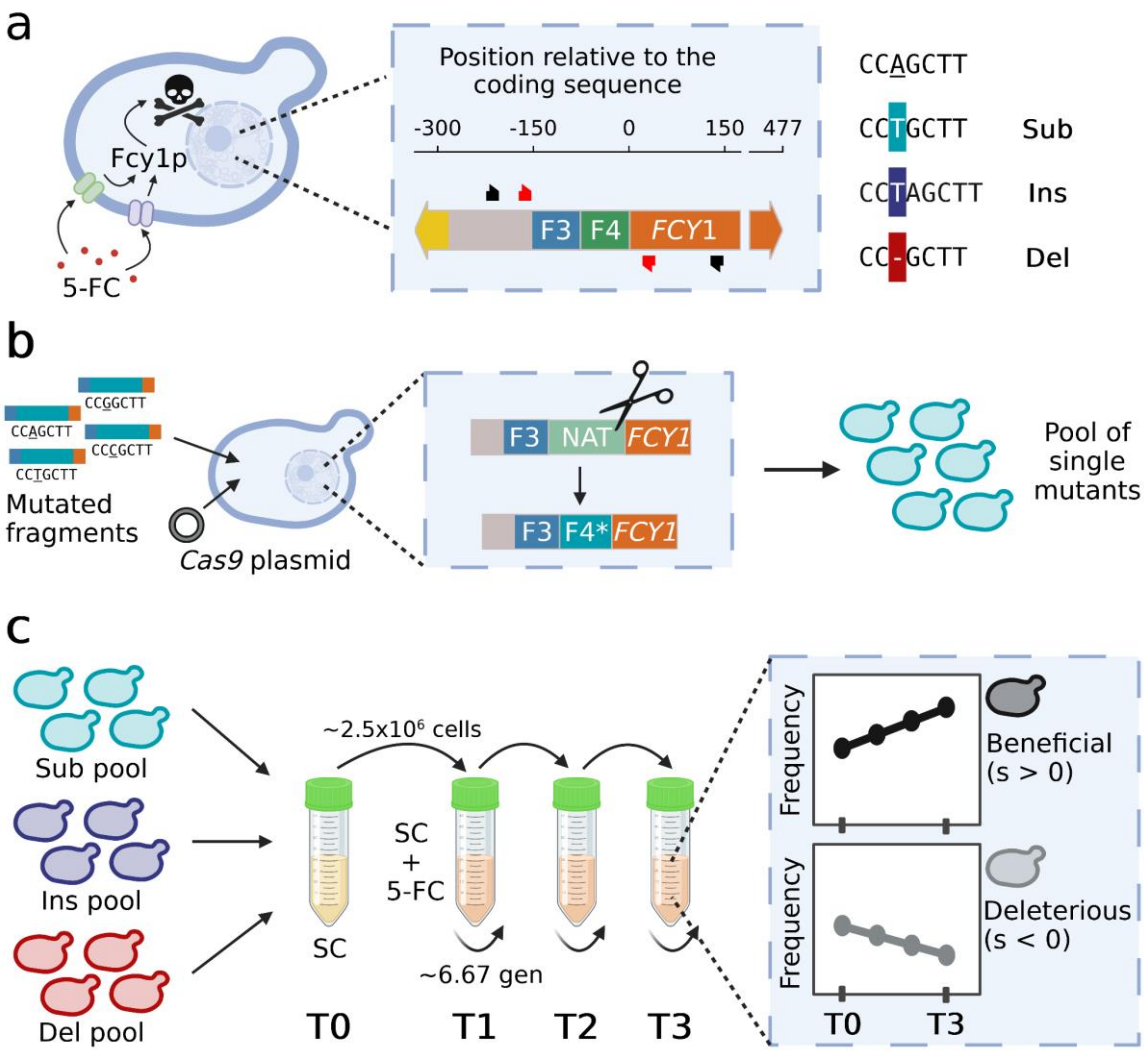


Fig. 2: Experimental design of the *FCY1* promoter deep mutational scanning. **a,** 5-FC is converted into 5-FU, a toxic compound, by the cytosine deaminase (Fcy1p), such that *FCY1* expression can have a strong effect on fitness. Fragments 3 (F3) and 4 (F4) of the *FCY1* promoter were replaced with a library of synthetic sequences corresponding to all possible single-nucleotide substitutions (Sub), insertions (Ins) and deletions (Del). The oligonucleotides used to sequence the region and genotype mutants are marked in red (short fragment) and black (extended fragment, including the recombination breakpoints). **b,** Mutated F3 and F4 sequences were inserted at the endogenous yeast *FCY1* locus by CRISPR-Cas9 genome editing. A NAT resistance cassette was first introduced in F3 or F4 by homologous recombination. Using a specific guide RNA, Cas9 could be targeted to the cassette, such that synthetic DNA fragments would be inserted at F3 or F4. This step was repeated four independent times. **c,** The three types of pools (Sub, Ins, Del) were combined and grown in synthetic complete media (SC) and then transferred to media with 5-FC to allow for competition based on the abundance of Fcy1p. Loss of function of *FCY1* has no detectable effect in SC media without 5-FC. At every timepoint, DNA was extracted and the edited locus was sequenced to follow the frequency of each mutant in the culture, allowing to estimate selection coefficient for each mutation from their enrichment or depletion through time. Schematics created in BioRender. Aubé, S. (2025) https://BioRender.com/frtbee2.

All 32 cultures were grown for ~20 generations in liquid media, with dilution every ~7 generations to maintain exponential growth (*Methods*). Samples were taken at each dilution step, as well as before the start of the experiment, resulting in a series of timepoints from T0 to T3 (**Fig. 2c**). The relative frequencies of all mutations were monitored by sequencing of the F3-F4 region (**Fig. 2a**). Due to the toxicity of Fcy1 activity in the presence of 5-FC, we expected that mutations reducing gene expression would prove adaptive and increase in frequency, outcompeting those resulting in a higher expression level. Accordingly, we estimated the selection coefficient (s) associated with each promoter sequence from the slope of its log-scaled frequencies relative to the WT[36] (**Fig. 2c**).

Mutant pools sampled from the competition experiment were first genotyped by sequencing a 174 bp fragment, which covered the full F3-F4 region and the first 20 bp of *FCY1* (red oligonucleotide pair; **Fig. 2a**). There were however major reproducibility issues. While estimated selection coefficients were consistent between culture replicates of the same transformation, they diverged substantially across transformation replicates (**Supplementary Fig. 1a**). In addition, comparisons of the distributions of selection coefficients across the different time intervals revealed a gradual shift of the mean. From ~0 for the T0-T1 interval, it decreased below -0.15 for the full competition experiment (**Supplementary Fig. 1b**). Through further experiments, we found that these effects were caused by secondary mutations introduced during library construction (**Supplementary Fig. 1c-d**). We modified our sequencing protocol to genotype a longer region (black oligonucleotide pair; **Fig. 2a**) and repeated the analyses to eliminate these effects. We explain the observations in detail in the supplementary material (**Supplementary Note**) as we suspect these could be common in genome editing studies and thus be of broad interest.

Fitness is robust to *FCY1* promoter mutations

After correcting for the impact of secondary mutations, the inferred selection coefficients for the F3-F4 mutants reveal only small fitness effects, which vary mostly uniformly along the mutagenized region (**Fig. 3a**). Across all assayed promoter mutations, median *s* estimates obtained from replicate experiments range from ~-0.037 to ~0.035 gen^{-1}, for a grand median of ~-0.002 gen^{-1}. Experimental noise moreover appears of a similar magnitude as these deviations from 0, as shown by an average difference of ~0.028 gen^{-1} between the highest and lowest *s* among replicate estimates for the same genotype, up to ~0.110 gen^{-1} (**Extended Data Fig. 4a**). A small number of secondary mutations remain undetected (not found within the longer sequencing fragment), resulting in outlier selection coefficients for one pair of culture replicates within some F3-F4 genotypes. Following stringent removal of these rare outliers (**Supplementary Fig. 8**), none of the 1020 *s* estimates for *FCY1* promoter mutations appear significantly different from 0 ($p > 0.05$; one-sample t-test with Benjamini-Hochberg correction at FDR=5%). Identifying significant effects using $|s|$ thresholds on the most extreme replicates instead results in small numbers of

significant changes (12 at $|s| > 0.01$), which are consistent with the null expectation from random permutations of the data (**Extended Data Fig. 4**). Overall, we thus conclude that there were no measurable fitness effects among the assayed single-nucleotide mutations.

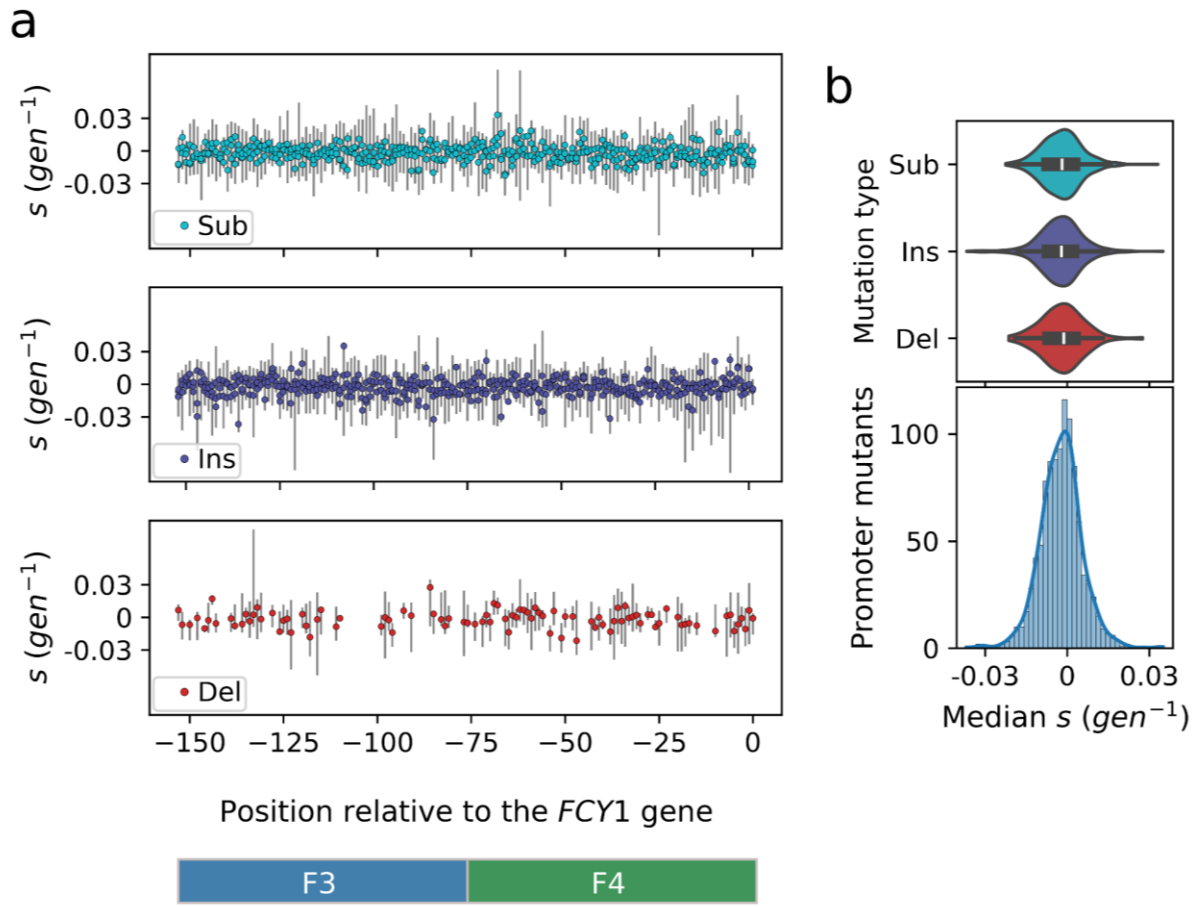


Fig. 3: Mutations in the *FCY1* promoter do not have a measurable impact on fitness in 5-FC conditions. a, Fitness effects of single-nucleotide mutations in the *FCY1* promoter. All substitutions (Sub; n=462), insertions (Ins; n=462) and deletions (Del; n=96) are shown. Points display the median selection coefficient for each mutant genotype, while error bars show the minimal and maximal values observed at each position along the mutated region. Nucleotide positions which appear to be missing are due to mutations which result in the same genotype (**Supplementary Fig. 2**), for instance deletion of any single base in a homopolymer region. Only data from two unbiased transformation replicates (each duplicated into two culture replicates; four samples in total) are shown (**Supplementary Note**, **Supplementary Fig. 6**). **b,** Distribution of mutational effects on fitness for single-nucleotide mutations in the *FCY1* promoter. The median selection coefficient obtained for each single mutant is shown, either as separate distributions for each of three classes of mutation (top) or as the full distribution (bottom; n=1020 observations, one per mutant).

This absence of mutational effects is not due to a failure of the competition experiment, as confirmed by two layers of controls. First, when considering all mutations (F3-F4 and secondary), the variance of estimated selection coefficients is indeed much

narrower for cultures without 5-FC (**Supplementary Fig. 9**). Second, selection coefficients inferred for the added coding sequence mutants correlate well with previously published variant effects[31] ($\rho = 0.780$ with $p < 10^{-64}$; **Extended Data Fig. 5a**). There is moreover almost perfect agreement between the two experiments regarding the number and identity of resistance-conferring mutations (**Extended Data Fig. 5b**). Together, these results confirm that adaptive changes could have been observed in our assay, if they could occur through single-nucleotide promoter mutations. We note that, whereas samples which contained coding sequence variants also included the complete pool of promoter mutations (**Extended Data Fig. 3**), the resulting cultures were not used to precisely estimate F3-F4 mutational effects. Although noisier, the corresponding selection coefficients were nonetheless qualitatively similar to those reported in **Fig. 3a** (**Supplementary Fig. 10**).

This lack of adaptive mutations within the *FCY1* promoter contrasts markedly with the high accessibility of adaptation through changes in the corresponding coding sequence. Within the *FCY1* gene, ~29% (408/1411) of single-nucleotide substitutions indeed result in codon changes which confer resistance to 5-FC in the same conditions (**Extended Data Fig. 5c**)[31]. Among these, only 67 introduce a premature stop codon. As such, although adaptation to 5-FC is readily available through *FCY1* mutations, it cannot be conferred by single-nucleotide changes in the promoter. In order to understand how this inaccessibility of adaptation arises, we first measured the effect of promoter mutations on *FCY1* expression.

Many mutations in the *FCY1* promoter substantially affect expression level

To obtain expression level measurements for all promoter mutants, we used a sort-seq approach[37,38]. The complete pool of F3-F4 single mutants was reconstructed in a genetic background where the distal part of the *FCY1* gene – starting from its 43rd base – was replaced with a mEGFP sequence (*Methods*, **Extended Data Fig. 6**). As such, the translation initiation context could be maintained while adding a fluorescent reporter to measure promoter activity. Yeast cells of the corresponding mutant library were sorted into five bins (numbered 0 to 4) according to their log-scaled fluorescence intensity (**Fig. 4a, Supplementary Fig. 11**). All the resulting samples were sequenced and read counts were used to compute a weighted average of bins for each genotype (*Methods*). The effect of each mutation was quantified using a Δ_{Ref} score equal to the difference between the corresponding mean bin and the average obtained for the WT. This was performed separately within each of six replicates (three transformation replicates, each subdivided into two culture replicates).

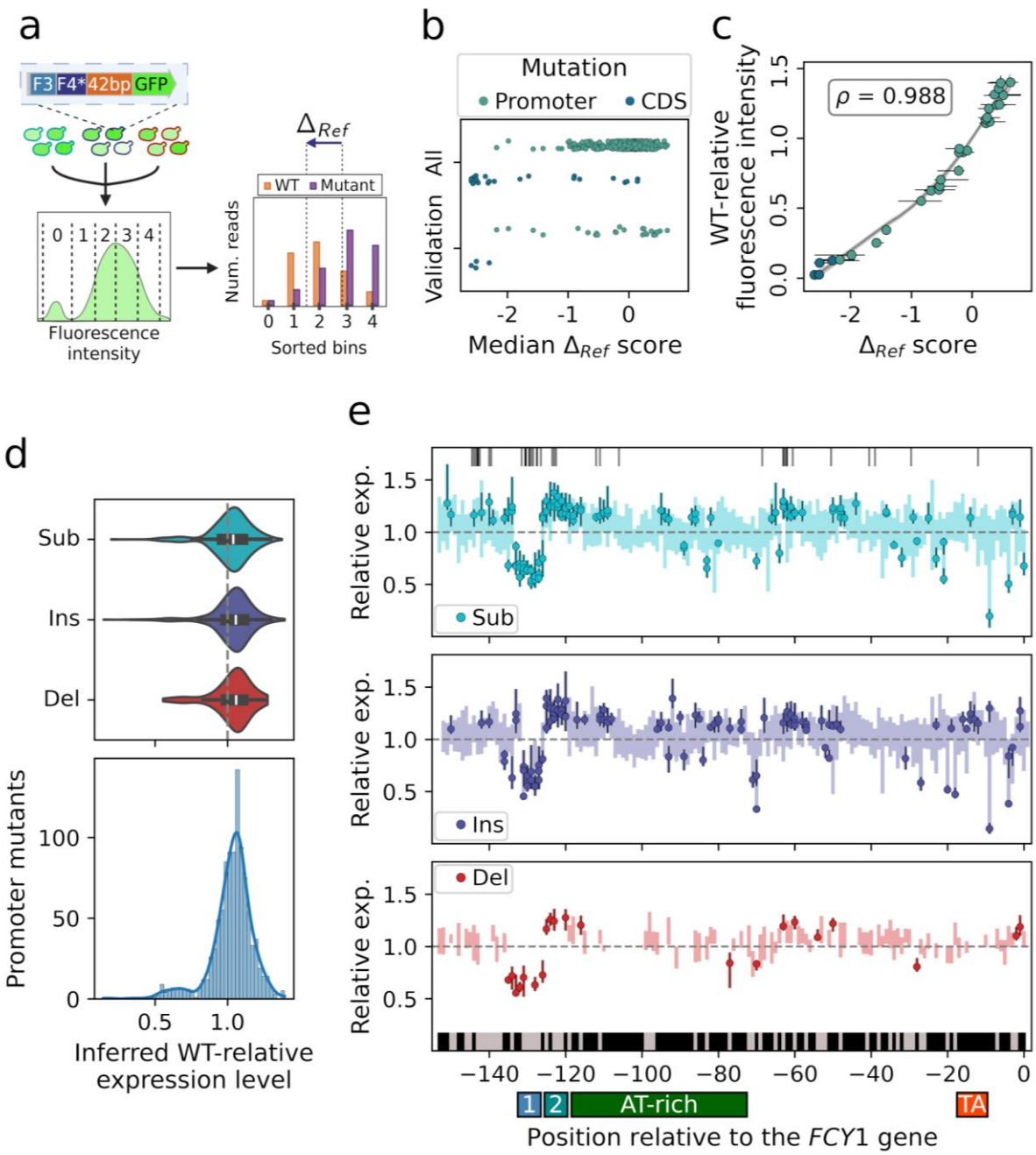


Fig. 4: Mutations in the *FYC1* promoter impact expression level. a, Schematics of the sort-seq experiment. Yeast cells from pools of promoter mutants were sorted into five bins according to their fluorescence intensity (representative distribution shown). From the sequencing of each bin, WT-relative fluorescence intensity could be estimated as a Δ_{Ref} score. Created in BioRender. Aubé,S. (2025) https://BioRender.com/uzzv5xy. **b,** Median sort-seq scores obtained for all variants detected across six replicate experiments. In addition to 1019 F3-F4 variants ("Promoter"), 28 single mutations in the first 42 bp of *FCY1* were identified ("CDS"), of which the vast majority are indels. The full distribution of scores ("All") as well as the 30 mutants selected for reconstruction ("Validation") are shown, split between the two categories in each case. **c,** Calibration curve of sort-seq scores**.** Points display the medians of sort-seq scores (n=6) and WT-relative fluorescence intensities (n=3) for all 30 validation genotypes. Horizontal and vertical error bars show the minimal and

maximal observations associated with each median. The high correlation between medians is shown with Spearman's rho ($p < 10^{-23}$). The line and the corresponding shaded area display a LOWESS fit with its 95% confidence interval (*Methods*). **d,** Distributions of WT-relative expression levels for all single-nucleotide promoter mutations, as inferred from the calibration of sort-seq scores. At the top, distributions are shown separately for substitutions (Sub; n=461), insertions (Ins; n=462) and deletions (Del; n=96), with a dashed line indicating WT expression level (1.0). At the bottom, the full distribution, combining all three types of single mutants, is shown (n=1019). **e,** Distribution of promoter mutations with a significant effect on expression (n=245), along the mutagenized region. Separately for Sub (n=113), Ins (n=111) and Del (n=21) variants, the estimated WT-relative expression (point) and its corresponding 95% prediction interval (error bar) are shown for each mutant affecting FCY1 expression by 5% or more (minimum or maximum of the corresponding 95% prediction interval; *Methods*). Shaded rectangles display the minima and maxima across all prediction intervals for each position along the sequence, including variants which are not shown. In the upper plot, vertical lines show the midpoint of hits (n=41) to known TF recognition motifs, identified using YeTFaSCo[39]. In the lower plot, the black and white rectangles respectively display A/T and G/C positions. Promoter positions corresponding with the two main clusters (sites "1" and "2"), the approximate AT-rich region and the TATA-like element ("TA")[40,41] are also identified.

A narrow range of scores was obtained, with the median Δ_{Ref} of most promoter mutations falling between -1 and 0.5 (**Fig. 4b**). Of the 1020 F3-F4 mutants, 1019 were covered in at least two replicates after abundance-based filtering (*Methods*). Many secondary mutations were also recovered (**Fig. 4b**), including indels in the first 42 bp of *FCY1*. Interestingly, while most such mutations were associated with a low fluorescence intensity and thus a negative Δ_{Ref}, some did not completely abolish expression. We specifically found out that indels which occurred before the second methionine codon of *FCY1* maintained ~10% of the WT fluorescence level (**Extended Data Fig. 7 a,b**). This is consistent with a previous report of a minor translation initiation site at this codon[42].

To validate the sort-seq assays, we selected 30 mutants for independent reconstruction (**Fig. 4b**). For each newly constructed strain as well as for the WT reference, fluorescence intensity was measured by flow cytometry (*Methods*). The resulting WT-relative intensities are in strong agreement with the corresponding median Δ_{Ref} scores (**Fig. 4c, Supplementary Fig. 12**). We additionally used this validation to draw up a calibration curve, enabling the conversion of sort-seq scores into more easily interpretable relative fluorescence intensities. To this end, we fit a LOWESS regression (**Fig. 4c**), to avoid making specific assumptions about the relationship between the two variables. This was performed within a bootstrapping routine which involved repeated resampling of Δ_{Ref} values as well as validation fluorescence measurements, thus generating 95% prediction intervals for the curve itself and for the relative fluorescence intensities of all genotypes (*Methods*, **Supplementary Table 9**). We hereafter refer to the medians of the latter intervals as

347 WT-relative expression levels for the corresponding mutations, where a value of 1
348 denotes the WT level.
349
350 Examining the inferred relative expression levels for all F3-F4 mutants reveals a slight
351 bias towards expression increases, with medians lying between 1.04 and 1.06 (**Fig.**
352 **4d**). That mutations are more likely to increase expression than to decrease it may
353 appear surprising, since Fcy1p is already highly expressed and is among the 10%
354 most abundant proteins in *S. cerevisiae*[43]. While this bias could be artefactual, the y-
355 intercept of the calibration curve suggests otherwise (**Fig. 4c**). It is very close to 1
356 (∼1.00, with 95% prediction interval from 0.98 to 1.02) – as expected for a Δ_{Ref} of 0 –
357 , although no WT measurements were used in the LOWESS fitting.
358
359 The overall distribution of effects highlights that most single mutations in the *FCY1*
360 promoter impact expression levels by less than 15%, with 75% of variants lying
361 between values of 0.98 and 1.11 (**Fig. 4d**). Yet, we also observe a small number of
362 mutations with larger effects, decreasing expression to ∼50% of its WT level, or even
363 lower. This tail appears more pronounced than previously reported in yeast[17,24],
364 potentially because we specifically mutated the most proximal part of the *FCY1*
365 promoter, where mutations are more likely to disrupt processes such as transcription
366 initiation. Among the 1019 F3-F4 mutations, 245 significantly affect *FCY1* expression,
367 by 5% or more (95% prediction intervals). In agreement with a previous study of a
368 yeast promoter[27], these significant mutations are not distributed uniformly along the
369 mutagenized promoter region (**Fig. 4e**). This distribution reveals important
370 determinants of *FCY1* expression level.
371
372 The fact that significant mutations form two main clusters, from positions -132 to -127
373 (site 1) and from -125 to -120 (site 2), first highlights the importance of transcription
374 factor (TF) binding. Within these two sites, 92.4 % (73/79) of all variants change
375 expression by 5% or more, in a consistent direction within each cluster. Both sites are
376 within a 80 bp segment known to be critical for TF binding in yeast promoters[18,37,44],
377 while the whole -132 to -120 region is a hotspot of predicted TF binding sites[39] (**Fig.**
378 **4e top; Supplementary Table 10**). As such, the clusters likely reveal TF binding sites.
379 Interestingly, the deletion of either of the two sites is not more impactful than the
380 strongest mutations within the corresponding cluster (**Extended Data Fig. 8**),
381 suggesting that single mutations are sufficient to disrupt them.
382
383 The position of the two most impactful mutations additionally highlights how alterations
384 of transcription initiation may severely reduce *FCY1* expression. These variants, a
385 substitution and an insertion which respectively decrease expression to ∼16% and
386 ∼13% of its WT level (cytometry; **Fig. 4c**), both occur at position -9 (**Fig. 4e**). This
387 position is immediately downstream of the previously identified TATA-like element of
388 the *FCY1* promoter[40,41] – from -17 to -10 –, where the transcription pre-initiation
389 complex assembles. Additionally, the transcription of *FCY1* has been shown to initiate

390 at position -7[33,45]. The corresponding mutational effects thus likely reflect the disruption
391 of transcription initiation. The conflicting report of a most abundant transcript isoform
392 beginning further upstream, at -29[46], however adds uncertainty.
393
394 Finally, the complete set of significant mutations underscores that most single-
395 nucleotide changes within the poly(dA:dT) region of the *FCY1* promoter weakly affect
396 its transcriptional activity. The corresponding AT-rich sequence, approximately
397 between positions -118 and -73 (17.39% GC; **Fig. 4e**, bottom), is indeed comparatively
398 depleted in strong mutational effects. Only 17.6 % (52/295) of variants within it are
399 significant, compared to 26.7 % (193/724) in the remainder of the mutagenized region.
400 This may seem contradictory, considering that such regions are important for the
401 establishment of nucleosome-depleted regions within promoters[47,48], which favor
402 transcription and can be sufficient to induce expression from a downstream core
403 promoter[49]. It is however consistent with previous work on the *URA3* promoter, which
404 showed that deletion of more than half of the corresponding 50 bp poly(dA:dT) was
405 necessary to abolish expression[50].
406
407 Overall, we show that 24% (245/1019) of single-nucleotide mutations in the *FCY1*
408 promoter have a significant effect on expression level, many of which appear to disrupt
409 the binding of TFs at two putative sites. As such, *FCY1* expression is sensitive to
410 mutations, meaning that the DME on expression level is not sufficient to explain the
411 absence of fitness changes in 5-FC conditions.
412

413 The expression-fitness relationship for *FCY1* is mostly flat

414 To investigate the mechanism underlying the absence of measurable fitness effects,
415 we set out to characterize experimentally the expression-fitness relationship for *FCY1*
416 (**Fig. 5a**). Six yeast promoters covering a wide range of expression levels were used
417 to construct two sets of strains, in which each promoter was inserted upstream of either
418 the native *FCY1* sequence or the previously described 42bp-mEGFP construct
419 (**Extended Data Fig. 6c**). The *FCY1* strains were grown in media with 5-FC (12 µM;
420 like in the competition experiment) – to obtain growth rate measurements –, while
421 mEGFP strains were used to measure expression levels (*Methods*). For each
422 promoter, the measurements from the two strains were combined to draw the
423 corresponding expression-fitness curve (**Fig. 5a**).
424
425 This experiment revealed a very steep fitness function, where growth rate decreases
426 sharply until *FCY1* expression reaches ~18% of its WT level and then plateaus (**Fig.
427 5b**). Consequently, this fitness curve is flat around the WT, such that small expression
428 changes have no fitness effect. The shape of this function is thus consistent with the
429 fact that no single-nucleotide promoter mutation conferred resistance to 5-FC in our
430 bulk competition experiment. To further validate this result, we leveraged previously
431 characterized validation mutants. The set of alleles used in the calibration of sort-seq
432 measurements, which also includes a small number of indels in the coding sequence,

433 was reconstructed in the native *FCY1* background and the resulting strains were
434 phenotyped in 5-FC conditions (*Methods*). In combination with the corresponding
435 relative fluorescence intensities, these growth measurements confirm the flatness of
436 the curve around WT expression level. Almost all promoter mutants display WT-like
437 growth rates, with no clear trend along the complete range of effects (**Fig. 5c**). Among
438 three assayed coding sequence mutations, the only one which reached the high
439 fitness region is associated with a ~98% decrease in fluorescence intensity
440 (**Supplementary Table 11**), caused by a frameshift-inducing indel. This highlights how
441 adaptation is only possible through very large reductions in Fcy1 abundance, which
442 cannot be achieved by single promoter mutations. Zooming in on the promoter
443 mutants further illustrates this point. Among them, the two single mutations with the
444 largest effect barely reach the threshold below which fitness rapidly increases,
445 although they respectively decrease *FCY1* expression to ∼16% and ∼13% of its WT
446 level (**Fig. 5d**). As such, the shape of the fitness function, jointly with the corresponding
447 DME, renders adaptation to 5-FC inaccessible from single changes in the *FCY1*
448 promoter.
449
450 The major contribution of the fitness function is made clearer by comparing the
451 selection coefficients inferred from the competition experiment to the distribution
452 expected from the relationship between 5-FC concentration and growth rate
453 (**Extended Data Fig. 2a**). Under the assumption that varying the abundance of Fcy1p
454 would have the same effect as changing the concentration of 5-FC (*Methods*), the
455 most impactful promoter mutation would have been associated with a *s* of 0.19, while
456 seven other mutations would have resulted in $s > 0.1$ (**Fig. 5e**). Adaptation would
457 then have been accessible through regulatory changes, albeit much less readily than
458 from mutations in the coding sequence.
459
460 Whereas growth rate varies according to 5-FC concentration, the shape of the fitness
461 function remains mostly constant. Across concentrations resulting in ∼10% to ∼70%
462 inhibition for the WT, growth rate variation is restricted to the same narrow range of
463 relative expression levels and only the height of the corresponding plateau varies
464 (**Extended Data Fig. 9a**). A similar phenomenon is observed in media lacking uracil,
465 where the deamination of cytosine into uracil by Fcy1p is essential for growth. With
466 increasing concentrations of cytosine, the height of the plateau of high fitness
467 increases, while the region of varying fitness remains mostly the same (**Extended**
468 **Data Fig. 9b**). At 5-FC and cytosine concentrations resulting in similar levels of relative
469 inhibition or growth, the fitness functions obtained in the two conditions are very
470 similar, with a sign inversion (**Extended Data Fig. 10ab**). This suggests that our
471 results also extend to the canonical cytosine deaminase function of the Fcy1 enzyme.
472 Growth measurements for a set of promoter mutants covering the full range of
473 mutational effects confirm that increasing *FCY1* expression, as allowed by the DME
474 of single mutations, is not adaptive under uracil starvation (**Extended Data Fig. 10c**).
475 Overall, these experiments demonstrate that adaptation is not accessible from single-
476 nucleotide changes in the *FCY1* promoter, in LOF (5-FC) as well as in gain-of-function

(cytosine) contexts. This is due in part to the small number of large mutational effects on gene expression, but more importantly to the shape of the fitness function, which is flat around WT expression level.

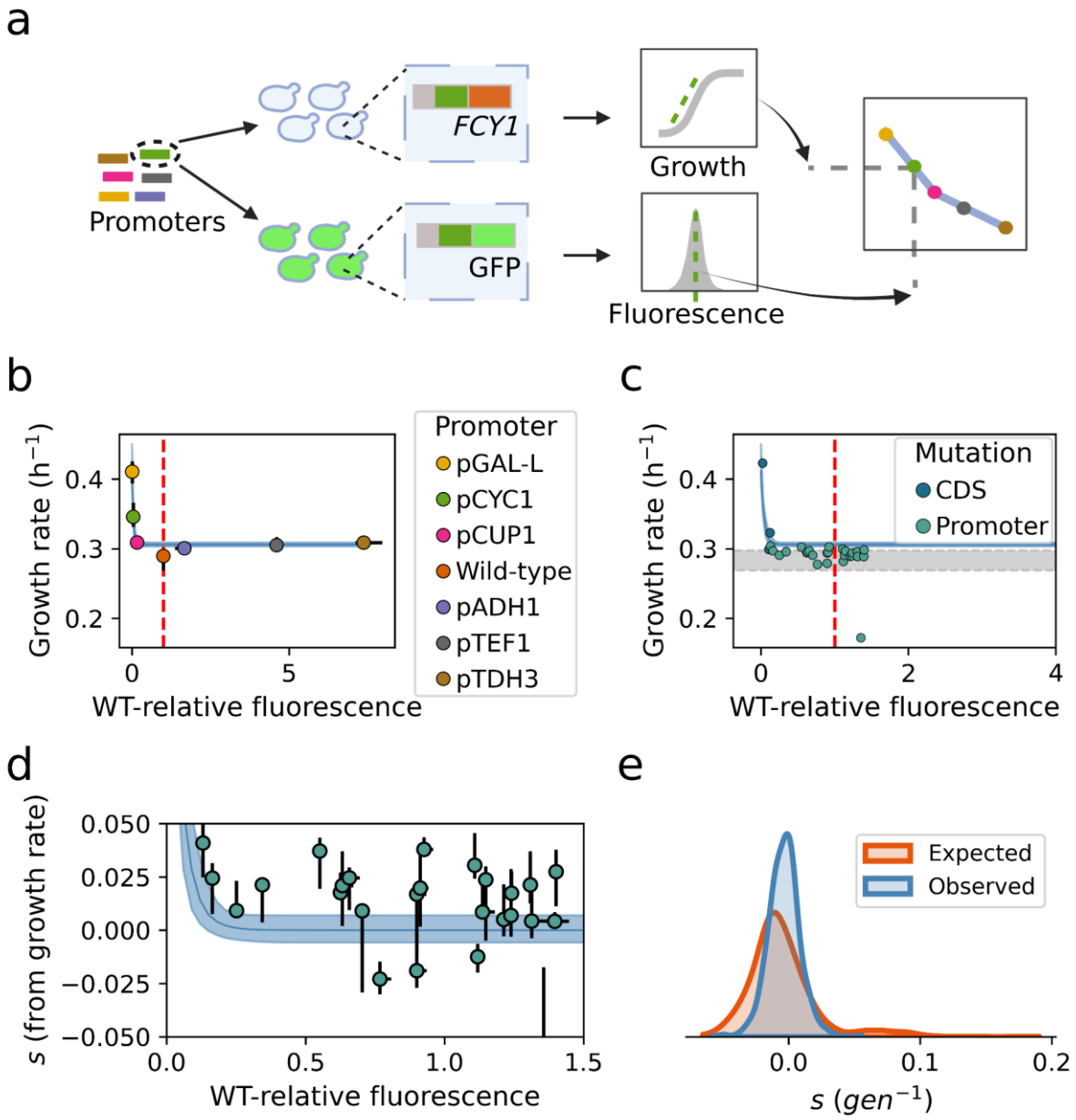


Fig. 5: The *FCY1* expression-fitness function is flat around WT expression level, making fitness robust to promoter mutations. a, Strains with varying expression levels were constructed to measure the fitness function. A set of yeast promoters covering a wide range of activities were introduced in the genome in place of the endogenous *FCY1* promoter, either upstream of *FCY1* or of a fluorescent reporter. Created in BioRender. Aubé, S. (2025) https://BioRender.com/1p60mp7. **b,** Expression-fitness relationship for *FCY1* in 5-FC conditions. All strains with *FCY1* expression equal to or higher than ~18% of its WT level display a growth rate which is indistinguishable from the WT. For each promoter, medians of growth rate and WT-relative fluorescence (n=3) are shown along with the corresponding minima and maxima (error bars). The fitness function is displayed as a fitted exponential decay function, with its corresponding 95% confidence interval (*Methods*). The WT strain,

which lacks a resistance cassette and is not directly comparable to the other ones (*Methods*), was not used in the fit. **c,** Growth rates of reconstructed mutants are consistent with the fitness function. For validation mutants (**Fig 4c**) which could also be reconstructed in the *FCY1* background, medians (n=3) are shown for growth rates in 5-FC [12μM] media and WT-relative fluorescence intensities. This includes single promoter mutations ("Promoter"; n=26) and indels in the first 42 bp of the gene ("CDS"; n=3). The dashed line identifies the WT expression level (1.0), while the shaded area displays the range of growth rates measured for the WT background (panel **a**), into which the mutations have been reconstructed. **d,** Reconstructed promoter mutations are mostly neutral. For the same set of single mutations in the promoter (n=26), medians (n=3) are shown for selection coefficients and relative fluorescence intensities, with error bars displaying the corresponding minimal and maximal replicates. Selection coefficients are computed from growth measurements (*Methods*) shown in the previous panel, relative to the WT strain. The previously fitted fitness function (panels **b** and **c**) is displayed with its corresponding 95% confidence interval, after conversion into selection coefficients. Only the region of the expression-fitness curve where promoter mutants lie is shown. The outlier mutant which is hidden is associated with a median *s* of ~-0.28. **e,** The shape of the fitness function prevents expression changes from affecting fitness. The DME on fitness observed in the competition experiment is much narrower than the one expected from the relationship between the concentration of 5-FC and growth rate (**Extended Data Fig. 2a**).

Discussion

The rapid accessibility of adaptive changes through regulatory mutations depends on their effect sizes on gene expression and on the relationship between fitness and expression level. One powerful approach to investigate these determinants is to focus on a gene where adaptation is known to arise and systematically assay all single regulatory mutations. To this end, we studied the promoter of the gene *FCY1*, a protein-coding region within which single-nucleotide changes have been shown to confer resistance to the drug 5-FC[31,51]. We show that no promoter mutation has a measurable fitness effect in yeast exposed to this molecule, although large changes in expression level (reaching a reduction of 87%) are observed. In this adaptive landscape, beneficial phenotypes are thus inaccessible from single nucleotide variation in the promoter, although they arise readily from mutations in the corresponding coding sequence[7,31].

Our initial data instead suggested that most promoter mutations were highly deleterious. Further investigations revealed that this result was due to secondary resistance-conferring mutations introduced during library construction, outside of the sequenced region (**Supplementary Note; Supplementary Fig. 1,4-8**). These beneficial mutations largely occurred in the WT promoter background, since this genotype was generated at high frequency within the mutant pool. This comparatively made all true promoter single mutants appear less fit. Interestingly, bias thus arose from the reconstruction of the WT within the CRISPR library, which is widely regarded as a best practice to ensure the accuracy of mutant fitness estimates[52]. Such artifacts

538 may be common in large-scale genome editing experiments, due to the increased
539 frequency of single-nucleotide mutations – and especially indels – in synthetic DNA
540 oligonucleotides[53]. This highlights the importance of designing library construction
541 workflows which minimize the use of genic regions as homology arms to guide
542 recombination. Alternatively, these segments should at least be fully included in the
543 genotyping region, as we did to correct our experiment.
544
545 Within our model system, the inaccessibility of adaptation through promoter mutations
546 emerges jointly from the DME on expression level and from the shape of the fitness
547 function. Since these two parameters vary across genes[1,17], the lower adaptive
548 potential of promoter mutations is only generalizable to those which are similar to
549 *FCY1* in this regard. There may be a substantial number of such genes, as the *FCY1*
550 promoter is a typical yeast promoter. It is TATA-less like the majority of such regions
551 [40,54], features the most common poly-T/poly-A configuration[48] and belongs to the most
552 abundant class of yeast promoters (2474/5378)[55]. Although this promoter is
553 bidirectional, hundreds of such regions have been identified in the *S. cerevisiae*
554 genome and it drives much higher expression of *FCY1* than of the opposite gene[56].
555 The endogenous expression level of Fcy1 is in addition not exceptional for an enzyme.
556 Among proteins included in the recent yeast9 genome-scale metabolic model of *S.*
557 *cerevisiae*[57], more than 16% (196/1161) are more abundant than Fcy1p[43]. Because of
558 these typical features, our results likely reflect general properties.
559
560 The generality of our findings is first shown by the distribution of expression changes
561 we report, which is consistent with previous work. The large-scale mutagenesis of
562 promoters and regulatory elements in yeast[17,24] and mammals[20–23] indeed showed that
563 only a minority of single mutations strongly affect expression. In the yeast *TDH3*
564 promoter, for instance, only 7.7% (18/235) of mutations were reported to decrease
565 expression by more than 6%[24]. Within the *FCY1* promoter, we instead observe that
566 17.2% of single-nucleotide changes (175/1019) reduce expression level by at least
567 this magnitude. As such, the rarity of strong expression-decreasing mutations is likely
568 not unique to our model. This is further underscored by genome-scale computational
569 predictions of cis-regulatory mutations in yeast, according to which 10% of mutations
570 account for 50% of expression variation[18]. These observations suggest that, rather
571 than being an exception, the DME on expression level which we report is probably
572 representative of many yeast genes. It is however worth noting that it is likely
573 incomplete, since only single-nucleotide indels were included. If indels of variable
574 lengths[58] had instead been considered, larger expression changes could potentially
575 have been observed.
576
577 Second, the shape of the fitness function, which severely restricts the adaptive
578 potential of promoter mutations, is not atypical. Saturating abundance-fitness curves
579 with a fitness plateau, like observed for Fcy1 (**Extended Data Figs 9, 10**), have
580 generally been reported for enzymes, across organisms[1,25,26,59–63]. They are
581 additionally expected from enzyme kinetics and metabolic control theory[60,64]. Whereas

582 the WT expression level can sometimes be close to the steep region where fitness
583 varies[1,63], it is more often away from this cliff, as we report here. In the case of *FCY1*,
584 the flatness of the fitness function around WT expression may arise from permeases
585 exerting control over metabolic flux, rather than the enzyme. This is suggested by the
586 observation that variations in 5-FC and cytosine concentrations have large effects on
587 fitness, contrary to *FCY1* expression changes (**Extended Data Figs 2, 9, 10**). A similar
588 phenomenon has been reported in bacteria, with the β-galactosidase enzyme and the
589 corresponding permease. While the former displays a step-like fitness function as
590 *FCY1*, the latter is associated with a much more curved fitness function, such that
591 small activity or abundance changes may have selective consequences[61]. More
592 generally, recent proteomics and metabolomics studies have shown that, in most
593 cases, changing the abundance of a single yeast enzyme has a weak impact on the
594 corresponding metabolic flux[57,65,66]. Overall, because the shape of the expression-
595 fitness relationship for *FCY1* likely emerges from fundamental principles, the
596 inaccessibility of adaptation through single promoter mutations may be generalizable
597 to a substantial number of genes, and especially enzymes.
598
599 Rapid adaptation through cis-regulatory changes may generally be rare, which would
600 be consistent with it being less accessible through mutation. For instance, increasing
601 the expression of drug target *ERG11* is a common mechanism by which fungi acquire
602 resistance to antifungals. Yet, it does not occur through single promoter mutations, as
603 only copy number variations (CNV) have been reported[67,68]. Similarly, whereas CNVs
604 of the sulfate transporter *SUL1* consistently arise in yeast evolved in sulfate-limited
605 conditions[69,70], single mutations in the corresponding promoter have been shown to
606 provide much less adaptive benefit[27]. There are however reports of single promoter
607 mutations conferring adaptation, especially in bacteria exposed to antibiotics[71–73].
608 Additionally, the analysis of natural variation between two yeast strains revealed an
609 enrichment of promoter mutations among variants with fitness effects[74]. In contrast,
610 mutations in noncoding regions have been found to be a very minor contributor to
611 adaptation in laboratory-evolved yeast[16]. Our results will help better understand in the
612 future why some genes are more likely to contribute to adaptation by regulatory
613 mutations by providing an integrative framework that connects mutations to phenotype
614 and fitness.
615
616 Contrary to promoter mutations, single-nucleotide substitutions in the *FCY1* coding
617 sequence readily confer adaptation to 5-FC[31] (**Extended Data Fig. 5**). In accordance
618 with the fitness function we report (**Fig. 5b**), this would mean that 29% (408/1411) of
619 coding mutations decrease the abundance or activity of the Fcy1 enzyme below 5%
620 of its WT level. As the strongest promoter mutation only decreases expression to 13%
621 of its WT level, this shows that, in *FCY1*, coding changes have larger molecular effects
622 than cis-regulatory ones. Whether this is generally the case across genes is unclear,
623 but saturation mutagenesis studies have shown that a substantial fraction of missense
624 mutations in protein sequences induce LOF[59,75–81]. Almost abolishing gene expression
625 may be comparatively difficult to achieve, as illustrated in yeast by genome-wide

626 spurious transcription[82] and by the expression of introduced exogenous and random
627 DNA[83,84]. As such, it is plausible that adaptation by LOF mostly occurs through coding
628 sequence mutations, as for *FCY1*. This may even be the case for rapid adaptation in
629 general, including gain-of-function. Two recent surveys of adaptive mutations, in yeast
630 evolved across different environments[16] and in pathogenic fungi exposed to diverse
631 antifungal drugs[7], have collectively identified only tens of adaptive noncoding changes,
632 compared with thousands of coding mutations. Yet, a large fraction of these mutations
633 are LOF-inducing. Elucidating whether coding or regulatory mutations generally
634 contribute more to adaptation will likely require additional experiments as well as a
635 systematic literature comparison similar to that of Hoekstra & Coyne[85].
636
637 These investigations should first focus on characterizing fitness functions. An
638 important role for fitness function shape in defining whether cis-regulatory or coding
639 mutations contribute more to adaptation was previously theorized from data on a
640 handful of mutations conferring β-lactam resistance in bacteria[86]. When the WT
641 resides on a fitness plateau, as is the case for *FCY1*, expression changes are unlikely
642 to provide a large fitness benefit (**Fig. 6a**). Coding mutations, on the other hand, may
643 cause complete LOF – like in our *FCY1* model – or bypass the fitness plateau by
644 modifying the curvature of the fitness function (in gain-of-function contexts), for
645 instance by changing the affinity of an enzyme for its substrate[86]. When the WT instead
646 resides in a region of the curve where fitness varies rapidly, expression variation
647 provided by regulatory mutations could be equally beneficial as coding changes (**Fig.**
648 **6b**), or even more if the fitness function is peaked and mutations may overshoot its
649 optimum (**Fig. 6c**). Second, a complementary avenue of investigation would be to
650 assess if the respective DME of promoter and coding regions are fundamentally
651 different at the molecular level. Within coding sequences, DME are often bimodal and
652 display peaks corresponding to WT-like and LOF mutations[59]. This is observed for
653 fitness effects[77,79–81] – as for the *FCY1* gene –, but also when other phenotypes like
654 fluorescence, protein activity and binding are measured[75,76,78,81,87,88]. Although such
655 multimodality can emerge from saturating abundance-fitness functions (like in **Fig.**
656 **5b**), it may also be a more general feature of mutations within coding sequences. This
657 could arise from nonlinear relationships between the biophysical properties of proteins
658 and fitness-relevant traits like protein abundance[59,87]. One example may be protein
659 stability, which dictates the fraction of correctly folded (functional) proteins through a
660 sigmoid function[79,89–91]. By potentially increasing the mutational accessibility of LOF,
661 this relationship could for instance underlie the major contribution of such mutations
662 to rapid adaptation[12,14,16]. Other properties of coding and cis-regulatory sequences
663 could make other categories of mutational effects more or less likely in one type of
664 region or the other.
665
666 Altogether, previous studies and our results demonstrate the importance of
667 considering the shape and curvature of fitness functions to understand the effects of
668 mutations and the molecular bases of adaptation. This goes beyond the relationship
669 between fitness and expression levels in microbes as it appears key for phenotypic

traits in general. For instance, it was recently shown that considering the full shape gene dosage response curves significantly enhances our understanding of genotype-phenotype relationships in humans[92].

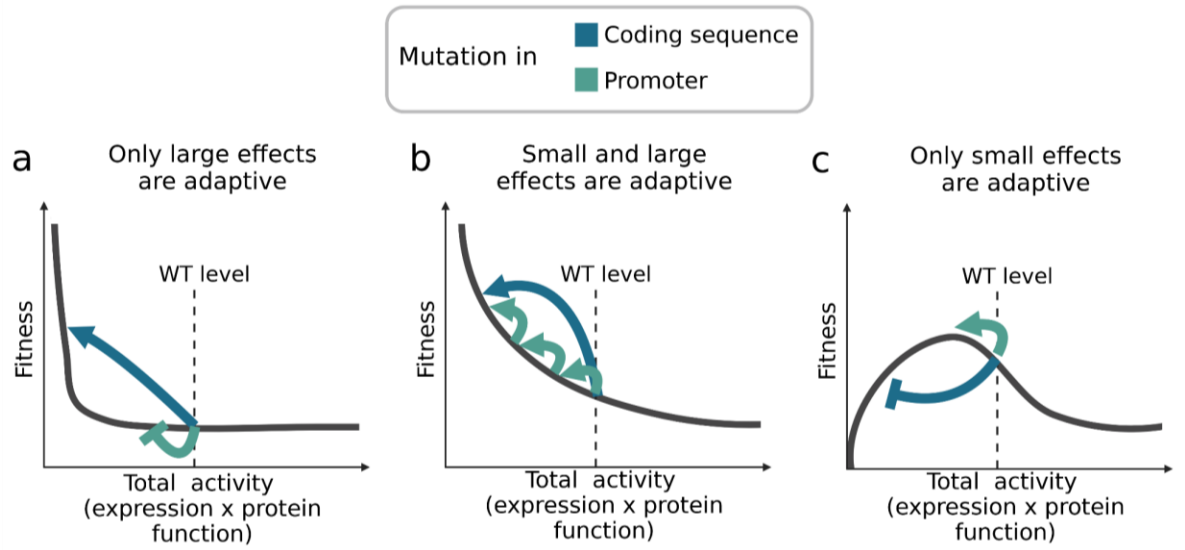


Fig. 6: Different shapes of fitness function may favor mutational effects of different magnitudes. Assuming that mutations in coding sequences have larger effects on molecular traits than changes within promoters, as is the case for *FCY1*, this can modulate whether adaptation is accessible through one type of mutation or the other. Although we only measured *FCY1* expression, we generalize expression-fitness relationships to total enzymatic activity. **a,** A step-like function, which is flat around WT, favors mutations with large effects on total activity. Accordingly, adaptation is accessible through coding sequence changes (arrow), while promoter mutations do not increase fitness and are not selected for (blunt arrow). This is what we report for *FCY1*. **b,** A fitness function with high curvature around WT allows mutations with small and large effects to be adaptive. Here, less impactful promoter mutations are also beneficial, such that adaptation can occur through one large-effect coding sequence mutation or through successive promoter mutations. **c,** A non-monotonic fitness function may favor small changes in total activity. Because they overshoot the curve's optimum, coding mutations with large effects are deleterious and thus not selected for (blunt arrow), while promoter changes with smaller effects can be adaptive (arrow). Created in BioRender. Aubé, S. (2025) https://BioRender.com/tj8z7wz.

Methods

Strains, oligonucleotides, plasmids and culture media

All *S. cerevisiae* strains constructed in this study were generated from a BY4742 (*his3Δ leu2Δ lys2Δ ura3Δ*) background[93] harboring a hygromycin resistance marker at the *HO* locus (*hoΔ::HPHMX*). This AKD1123 parental background and all other strains are listed in **Supplementary Table 1**, along with the identity of the corresponding genomic inserts and PCR primers as well as the selective media used. The sequences and references for all oligonucleotides used in this study can be found in **Supplementary Table 2**, while references for all plasmids are listed in **Supplementary Table 3**. Recipes for culture media are detailed

703 in **Supplementary Table 4**. When specified, selective media was prepared by adding one or
704 multiple of the following agents: geneticin [200 µg/mL] (G418; Bioshop Canada),
705 nourseothricin [100 µg/mL] (NAT; Jena Bioscience) and hygromycin B [250 µg/mL] (HYG;
706 Bioshop Canada). All Sanger sequencing was performed at the sequencing platform of Centre
707 de recherche du CHU de Québec-Université Laval (Université Laval, Québec, Canada).
708

709 ## Yeast transformation

710 Whenever inserts which included a selection marker had to be inserted in the yeast genome,
711 a standard lithium acetate yeast transformation protocol was used[94]. After selection of
712 transformants on the appropriate selective media, four to eight clones were selected for
713 validation PCR, depending on the strain. Unless stated otherwise, colony PCR was performed
714 on each side of the integration site to confirm proper integration, using Taq polymerase
715 (Bioshop Canada). All genomic inserts were amplified using the KAPA HiFi HotStart DNA
716 polymerase (Roche Molecular Systems).
717
718 This approach was for instance used to delete each fragment F1 to F4 of the intergenic region
719 between *JID1* (*YPR061C*) and *FCY1* (*YPR062W*), as well as the full intergenic segment and
720 the complete *FCY1* region (including the corresponding gene), by the insertion of a *natNT2*
721 cassette[95]. This resulted in a set of six strains harboring a different deletion of the *FCY1* region
722 (SA0127, SA0128, SA0129, SA0130, SA0131, SA0132). Similar deletions of regions F3 and
723 F4 were also performed in the 42bp-mEGFP construct obtained by CRISPR-Cas9 yeast
724 transformation (see below), generating fluorescent strains SA0246, SA0247 and SA0254
725

726 ## CRISPR-Cas9 yeast transformation

727 When sequences needed to be inserted in the yeast genome without any flanking selection
728 marker, a CRISPR-Cas9 approach was used. This required the prior generation of a strain
729 harboring a *natNT2* cassette at the chosen insertion site, which was done through standard
730 yeast transformation as described above. Competent cells from this genetic background could
731 then be cotransformed with the pCAS-NAT plasmid (**Supplementary Table 3**) – encoding a
732 guide RNA engineered to target the Cas9 endonuclease to the *natNT2* cassette[34] – and the
733 desired repair DNA, amplified with KAPA HiFi HotStart DNA polymerase (Roche Molecular
734 Systems). A previously published CRISPR-Cas9 transformation protocol was followed[96].
735 Briefly, competent cells were prepared as described, and transformed with unpurified PCR
736 amplification of the insert, DNA carrier [10 mg/ml], pCAS-NAT and PLATE solution
737 (polyethylene glycol 2000 [40%], lithium acetate [0.1 M], Tris-EDTA [10 mM Tris-Cl, 1 mM
738 EDTA] buffer [pH=8.0]). After mixing, cells were incubated for 30 minutes at 30°C, heat
739 shocked for 15 minutes at 42°C, resuspended in YPD media and incubated for 4-6 hours at
740 30°C. Transformation reactions were plated on YPD media containing G418, to select for
741 uptake of the pCas plasmid, and incubated at 30°C for 48 hours. Eight colonies from each
742 transformation were then grown overnight (30°C with shaking) in liquid YPD and streaked on
743 solid YPD. After 48 hours of incubation at 30°C, one isolated clone from each streaking was
744 grown overnight in YPD (30°C with shaking) and spotted on both YPD and YPD+G418 solid
745 media, to confirm the loss of the pCAS plasmid. Because the AKD1123 parental background
746 is resistant to hygromycin, HYG was added to the YPD throughout. Correct integration was
747 then validated by colony PCR and Sanger sequencing.
748

749 Using this protocol, the reinsertion of deleted regions F2, F4 and F1-F4 (full intergenic) was
750 performed, resulting in strains SA0162, SA0166 and SA0170. It was also used to generate
751 the 42bp-mEGFP construct (SA0228) to measure *FCY1* expression level and all validation
752 clones expressing either *FCY1* or mEGFP from the *FCY1* locus.
753

754 ## Growth assays

755 Unless stated otherwise, the following protocol was used each time the growth of *FCY1* strains
756 was assessed in media containing 5-FC (Alfa Aesar) or cytosine (Alfa Aesar). Each yeast
757 strain of interest was first inoculated in three replicates in SC [pH=6] liquid media – potentially
758 with HYG, NAT and/or G418, depending on each strain's resistance marker(s) – and grown
759 overnight at 30°C with shaking. After ~16 hours of incubation, a sample (100 µl) of each culture
760 was centrifuged and resuspended in the same volume of sterile water, before being aliquoted
761 and diluted again in sterile water to obtain up to 250 µl of cells at 1 optical density (OD) unit/ml.
762 Growth assays were then assembled in Greiner flat-bottom transparent 96-well plates as
763 follows: 160 µl of media (SC complete for 5-FC or SC-Ura for cytosine); 20 µl of diluted cells
764 [OD=1]; 20 µl of 5-FC or cytosine diluted to obtain the desired final concentration in a volume
765 of 200 µl. Each plate was incubated for 48 hours at 30°C without shaking in a Tecan Infinite
766 M Nano (Tecan Life Sciences) and OD_{595} was measured at 15 minutes intervals.
767
768 To obtain exponential growth rates, the exponential part of each growth curve was first
769 identified. The end T_{max} of the exponential phase was defined as the time where the derivative
770 of OD, estimated as the moving median of differences between consecutive OD
771 measurements, reached its maximum. To limit experimental noise, the first two hours of OD
772 measurements were discarded, while any T_{max} above 30 hours or below 3 hours was ignored.
773 In such cases, the first ten hours of the growth curve were instead defined as the exponential
774 phase. Exponential growth rates were then estimated as the slope of the linear regression of
775 $log(OD)$ over time. When converting the growth rate of any strain i into selection coefficient
776 s_i, the following formula was used[97]:

$$s_i = \frac{\mu_i - \mu_{WT}}{\mu_{WT}} \log 2$$

777
778 Variables μ_i and μ_{WT} are the growth rates for the strain of interest and the corresponding WT
779 reference, respectively. This comparison was always made using WT measurements obtained
780 in the same experiment as the growth rate of strain i.
781

782 ## Construction of the mutant libraries

783 ### Amplification of mutated F3 and F4 promoter fragments

784 All single-nucleotide substitutions, insertions and deletions in the F3 and F4 regions were
785 ordered as six separate oPools Oligo Pools (Integrated DNA Technologies Inc.) containing
786 sequences of 117 bp, 118 bp and 116 bp, respectively. Within the different pools, each
787 oligonucleotide is a variant of the full-length F3 or F4 region (77 bp) with flanking 20 bp
788 homology regions at both ends. The latter sequences are respectively the 20 bp immediately
789 upstream or downstream of the corresponding fragment in the reference genome of *S.*
790 *cerevisiae* S288C, as retrieved from the Saccharomyces Genome Database[98]. For
791 substitutions, each position of each fragment was iteratively degenerated into all four possible
792 nucleotides by specifying a N base, while for insertions, the same mixed base was inserted in
793 3' of each position in the reference F3 and F4 sequences. When such an N is encountered,
794 an equal ratio of each of the four standard nucleotides (A, T, G, C) is dispensed during

795 synthesis. For the two pools of deletions, single-nucleotide removal was performed iteratively
796 at each position of F3 and F4 and duplicates (occurring whenever two or more consecutive
797 nucleotides are identical in the reference) were removed to obtain non-redundant sets. All
798 sequences ordered as part of the six oPools are listed in **Supplementary Table 5**.
799
800 Upon reception, each oPool was spun down and resuspended in fresh PCR-grade water, to
801 obtain a final concentration of 50 µM. Aliquots of 100 µl at 0.05 µM were then obtained by
802 serial dilution from 2 µl of the initial resuspension. From these aliquots, PCR amplification of
803 each oPool was performed in four replicates using KAPA HiFi HotStart DNA polymerase
804 (Roche Molecular Systems) and the following cycle: 98°C for 3 min; {98°C for 30 s; 60°C for
805 15 s; 72°C for 10 s} x 22; 72°C for 20 s; Hold at 10°C. Each reaction was done in 50 µl:

Volume (µl)	Reagent
33	PCR-grade water
10	KAPA HF buffer [5X]
1.5	dNTPs [10 mM]
1.5	Forward oligo [10 µM]
1.5	Reverse oligo [10 µM]
1.0	KAPA Hi-Fi HotStart DNA polymerase
1.5*	oPool [0.05 µM] (or PCR-grade water)
50	**Total**

806 In each case, the diluted oPool resuspension was not added to the reaction mix and was
807 instead added individually in each reaction tube. In addition, one control without the oPool
808 dilution was performed for each primer combination. Both primer pairs were designed to
809 extend the 20 bp homology regions included in the oPool, such that mutated F3 and F4
810 featured 60 bp homology arms upstream and downstream of the corresponding genomic
811 region. The two pairs were CLOP299-C3 / CLOP299-C4 and CLOP299-D3 / CLOP299-D4,
812 respectively to amplify F3 and F4 oPools.
813
814 **Genomic insertion of mutated fragments in *S. cerevisiae***
815 To insert promoter mutations upstream of *FCY1*, competent yeast cells harboring a *natNT2*
816 deletion of the F3 or F4 region (strains SA0129 and SA0130) were transformed with the
817 corresponding PCR-amplified oPools. For each mutation category and each region, four
818 replicate transformations were performed from the corresponding four PCR reactions
819 described above – for a total of 24 transformations. The protocol described by Ryan *et al.*
820 (2016)[96] was followed (see *CRISPR-Cas9 yeast transformation*, above). After incubation,
821 G418-positive colonies from each agar plate were collected separately. Liquid YPD (5 ml) was
822 added to the plate, cells were resuspended by scraping with a glass rake and the suspension
823 was transferred to a 15 ml Falcon tube. From this suspension, glycerol stocks (687 µl cells
824 mixed with 313 µl sterile [80%] glycerol) were made for each transformation and stored at -
825 80°C. Two 500 µl aliquots from each tube were spun down and stored at -20°C after aspiration
826 of the supernatant. From these samples, mutant libraries were validated by PE150 sequencing
827 on an Illumina MiSeq instrument (Genomics Analysis Platform [PAG] of Institut de Biologie

828 Intégrative et des Systèmes, Université Laval, Québec). This approach was also employed to
829 insert the same set of promoter mutations upstream of the 42bp-mEGFP construct, using
830 strains SA0246 and SA0254. In this case, three replicate transformations were instead
831 performed per amplified oPool.
832
833 **Selection of coding sequence mutations**
834 Fifteen codon positions of the *FCY1* gene, for which all possible states had previously been
835 generated by DMS[31], were selected to obtain a diversity similar to that of the complete
836 mutational scanning of the F3-F4 region (15 positions x 64 codons = 960 mutants). Among
837 these positions, six were chosen for their wide range of variant effects and nine others were
838 selected randomly, to fully cover the previously reported range of mutational effects. The 15
839 corresponding pools of mutants were each grown for ~24 hours in 4 ml SC complete media
840 (30°C with shaking), from 25 µl of glycerol stock. Optical densities were measured and all
841 cultures were mixed in equal amounts to make four replicate 1 ml glycerol stocks (687 µl mixed
842 cultures and 313 µl sterile [80%] glycerol) of this pool of Fcy1p mutants, which were stored at
843 -80°C.
844
845 **Pooling of mutants**
846 The different sets of promoter mutations (substitutions, insertions and deletions, separately
847 for F3 and F4) were grown separately in 4 ml SC complete media, from 50-100 µl of the
848 corresponding glycerol stock. After ~16 hours of incubation at 30°C with shaking, OD_{595} was
849 measured for all cultures. Subpools were then mixed to obtain equal proportions of all single
850 mutants, according to their OD_{595} and how many unique genotypes they each contained.
851
852 For mutations in the *FCY1* background, the four replicate pools of coding sequence mutations
853 were also grown from 75 µl of glycerol stock. All subpools were then combined as described,
854 according to their replicate number (i.e. replicate 1 for F3 substitutions was mixed with
855 replicate 1 of all other subpools). Each subpool of promoter mutations was thus included in
856 only two final pools: one with added coding sequence mutations and the other without. For
857 mutations in the 42bp-mEGFP background, all subpools of promoter variants were combined
858 as described without the addition of other mutants, except that there were only three replicates
859 of each.
860
861 **Bulk competition experiment in 5-FC media**
862 Each mutant pool was grown in two replicates in 4 ml SC complete media (30°C, 200 rpm) for
863 ~20 hours, from 111 µl or 65 µl of the corresponding glycerol stock. These inoculum volumes
864 were chosen to have at least 2000x cells for each mutant, respectively for pools with and
865 without added coding sequence mutations. The competition experiment was initiated by
866 diluting each preculture at final OD_{595} of 0.01 in 25 ml of SC complete media (with and without
867 5-FC [12 µM]) in 50 ml Falcon tubes. T0 samples were also prepared by centrifuging two
868 microtubes containing 5 OD_{595} units for each preculture (850 g for 5 minutes; 4°C). After
869 aspiration of supernatant, cell pellets were stored at -20°C. Newly inoculated cultures were
870 incubated (30°C with shaking) in the dark until OD_{595} reached ~1.0. At this first timepoint T1,
871 each culture was diluted in at final OD_{595} of 0.01 in 25 ml of fresh media. Additional cell pellets
872 were prepared from the remaining cultures. The full volume was first centrifuged at 4°C (1000
873 g; 5 minutes). Cell pellets were resuspended in ~5 ml of supernatants and two samples of ~5
874 OD_{595} units were prepared as previously. This process was repeated at timepoint T2 when the

875 new cultures reached OD_{595} ~1.0 and final T3 samples were prepared when the newly
876 inoculated T2-T3 cultures reached the same OD_{595}. The timings of each timepoint and the
877 corresponding OD_{595} values for all mutant pools are listed in **Supplementary Table 6**.
878

879 ## Sort-seq experiment

880 One final pool of F3-F4 mutants in the 42bp-mEGFP background was grown overnight from
881 100 µl glycerol stock in 5 ml of SC complete [pH = 6.0] media, at 30°C with shaking. After
882 measurement of its OD_{595}, the culture was diluted to 0.15 OD/ml in 20 ml of fresh SC complete
883 [pH = 6.0] and incubated for 5 hours at 30°C with shaking, to reach exponential phase (~0.5
884 OD/mL). From this, 4.5 ml were transferred to a 5 ml polystyrene round-bottom Falcon tube.
885 The remaining 15.5 ml were centrifuged to collect cells, so that genotype frequencies before
886 sorting could also be measured by sequencing.
887
888 A FACSMelody cell sorter (BD Biosciences) was used to sort yeast cells based on their
889 fluorescence intensity. Five bins were defined along the distribution of log_{10} fluorescence
890 intensities, numbered 0 to 4 from the lowest to the highest fluorescence levels (Supplementary
891 Methods). More cells were collected from the most populated parts of the distribution (50 000
892 cells each from bins 2 and 3, versus 10 000 cells each from bins 0,1 and 4). Each bin was
893 sorted twice, resulting in two technical replicates, in a 96 well plate. The resulting fractions
894 were each used to inoculate 3 ml of SC complete [pH = 6.0] media in a 24 deep-well plate and
895 grown for 24h at 30°C with shaking (400 rpm). Each culture was spun down at 3000 rpm (1000
896 g) for 5 min and all pellets were washed with sterile water and transferred to a 1.5 ml
897 microtube. All microtubes were centrifuged at 15 000 rpm (21 000 g) for 1 min, supernatants
898 were removed and pellets were frozen at -80°C. This process was repeated two times for the
899 two other biological replicates, such that six replicate samples were obtained for each bin.
900

901 ## DNA library preparation and sequencing

902 The same approach was used for all sequencing libraries, for both the bulk competition and
903 sort-seq experiments. Genomic DNA was first extracted from each cell pellet using a standard
904 phenol-chloroform extraction protocol[99], including RNAse treatment. For the resequencing of
905 bulk competition samples in PE300, an additional RNAse treatment (10 minutes, 37°C) was
906 performed on previously purified DNA samples, followed by precipitation with sodium acetate
907 and cold ethanol. A first PCR was then performed using Q5 High-Fidelity DNA Polymerase
908 (New England Biolabs) to amplify the genotyping region chosen for each type of samples,
909 which included the F3-F4 promoter regions and flanking sequences of varying lengths. This
910 PCR step added Illumina primer-binding sites at both ends, separated from the fragment of
911 interest by small numbers of N bases (between 2 and 8; 10 total per PCR amplicon). For the
912 initial (PE150) sequencing of bulk competition samples, primers CLOP311-C6 to CLOP314-
913 H9 were used (as ordered in **Supplementary Table 2**). Two reactions were made from each
914 DNA extraction from cultures with added coding sequence mutants. One was specific to the
915 F3-F4 pool while the other targeted only the added *FCY1* variants. This was possible because
916 the pool of coding sequence mutations was generated from a codon-optimized sequence,
917 harboring many nucleotide changes from the WT [31]. For the PE300 resequencing of the same
918 samples, only F3-F4 PCR were performed, using oligos CLOP328-A6 to CLOP328-H7.
919 Finally, for the sequencing of sort-seq bins, also in PE300, oligos CLOP328-A6 to CLOP341-
920 A10 were employed. In this case, 7 N bases were added per amplicon.
921

922 In all cases, amplicons from this first PCR step were purified with AMPure XP magnetic beads
923 (Beckman Coulter Life Sciences) and diluted to 0.01 ng/µl. These dilutions were used as
924 template for a second set of PCR reactions using Q5 polymerase, which added Nextera
925 barcodes and Illumina adapters i5 and i7. A unique combination of i5 and i7 indexes was used
926 for each DNA sample. The final PCR products were also purified with AMPure XP beads,
927 before quantification using a NanoDrop spectrophotometer (Thermo Fisher Scientific).
928 Amplicons from the same sequencing round (PE150 competition, PE300 competition or
929 PE300 sort-seq) were mixed in equal proportions based on the NanoDrop quantification and
930 the resulting pool was sent for sequencing. PE150 sequencing was performed on a NovaSeq
931 Illumina instrument (sequencing platform of Centre de recherche du CHU de Québec-
932 Université Laval, Québec), while PE300 sequencing was performed on an AVITI instrument
933 from Element Biosciences (Genomics Analysis Platform [PAG] of Institut de Biologie
934 Intégrative et des Systèmes, Université Laval, Québec).
935

936 ## Analysis of sequencing reads

937 Following each sequencing run, reads were received already demultiplexed, according to the
938 sample id associated with each combination of i5 and i7 adaptors used. Read quality was first
939 assessed using FastQC (v 0.12.1) and MultiQC (v 1.12)[100]. Paired-end reads were next
940 merged using PANDAseq (v 2.11)[101], with the '-k' parameter set to 4. The N positions which
941 had been added at both ends during the first PCR step of library construction were trimmed
942 using command "fastx_filter" of VSEARCH (v 2.28.1)[102], according to a table listing the number
943 of N added for each sample. The first and last 20 bp of the resulting sequences, which were
944 the binding sites for the oligonucleotide primers used in the first PCR step, were also trimmed
945 using the same command. Following these trimming steps, all identical sequences were
946 aggregated with the "derep_fulllength" command of VSEARCH. Each unique sequence was
947 aligned to the corresponding reference (depending on the type of sample and the length of
948 the sequenced region) using the needle aligner from the EMBOSS suite (v 6.5.7)[103]. For
949 promoter (F3-F4) libraries, the opening and extension of gaps was equally penalized (4,4),
950 while penalties of 50 and 0.5 were respectively used for opening and extension for *FCY1*
951 coding sequence libraries[31]. Custom Python functions were then used to call variants from
952 each alignment. In addition to WT reads, only sequences harboring single-nucleotide
953 mutations (F3-F4 libraries) or single-codon substitutions (*FCY1* libraries) were kept for
954 downstream analysis. These variants were further filtered by length, to remove those that
955 differed too much from the expected length of the PCR fragment. For bulk competition
956 samples, read counts of each single mutant were converted to WT-relative frequencies. These
957 read counts were instead used as is in the downstream analysis of sort-seq samples.
958

959 ## Estimation of fitness effects

960 ### Calculation of selection coefficients

961 For each bulk competition culture, time series of WT-relative genotype frequencies were
962 reconstituted from variant frequencies obtained at the four timepoints (T0 to T3). The number
963 of generations associated with each timepoint was obtained from the corresponding durations
964 in hours, which varied between cultures, and the doubling time of the WT strain in the same
965 conditions (SC media with or without 5-FC [12 µM]), as estimated from growth curve
966 experiments. For each single-mutant genotype, the selection coefficient s was obtained by
967 fitting a linear model using the number of generations and the log-transformed WT-relative
968 frequencies as dependent and independent variables, respectively. This was performed

through ordinary least squares regression, using the "ols" implementation of the statsmodels (v 0.12.2)[104] Python module, with an intercept added to the model. Regressions were performed iteratively over all time intervals which included the T0 (T0-T1, T0-T2, T0-T3; with intermediate data points when applicable). Because selection coefficients were inferred separately within each culture replicate, as many replicate values as there were cultures were obtained for each of the intervals mentioned above.

To identify the selection coefficients which should be used in further analysis, two diagnostics were performed. First, the agreement between the initial (T0) log-scaled WT-relative frequencies and the fitted intercept from each regression model was assessed. Second, the linearity of log-scaled frequencies through time, as expected for cultures growing exponentially, was validated graphically. This revealed that no interval was reliable for the initial 150 paired-end genotyping of promoter (F3-F4) libraries, while T0-T2 selection coefficients should be used for the 300 paired-end resequencing of the promoter fragments and the 150 paired-end sequence of *FCY1* coding sequence (see Supplementary Note). The chosen selection coefficients were then used to identify a T0 abundance threshold for variant filtering, such that inferred s would not be correlated with initial read counts. This threshold was set at 85 reads, both for the promoter (300 paired-end) and coding sequence (150 paired-end) sequencing libraries. To ensure that low frequency but beneficial secondary mutations could be assessed, any variant which was below the threshold at T0 but over it at the end of the experiment was kept.

Removal of outliers and aggregation of replicates

In 5-FC samples, mutations sometimes appeared beneficial in only one pair of cultures (coming from the same transformation replicate), likely due to secondary resistance-conferring mutations (See Supplementary Note). Outlier removal was thus performed, separately for two types of libraries (promoter or coding sequence). The promoter (F3-F4) libraries were further separated depending on whether coding-sequence mutants had been added in the corresponding cultures. Within each group of samples, the median absolute deviation (MAD) of selection coefficients was first computed for each mutant. This was then used to compute robust z-scores for each replicate measurement within each genotype. The resulting distribution of robust z-scores calculated across all mutant genotypes was compared and its 95th percentile was defined as the threshold to identify outlier replicates. The filtering of outliers was however restricted to a subset of mutants which displayed a higher variance, to ensure that small absolute variations between replicates were not identified as outliers. This was assessed from each genotype's standard deviation, with the 95th percentile of the overall distribution used as threshold. For F3-F4 sequencing libraries obtained from cultures with added coding sequence mutations, the 85th percentile was instead used.

To obtain final estimates of the fitness effects of promoter mutations, only F3-F4 libraries without added coding sequence mutants were taken into account. Four of the eight replicate cultures, corresponding to two of the four transformation replicates, were additionally ignored. In these four samples, selection coefficients displayed a temporal shift towards negative values which appeared consistent with some undetected resistance-conferring mutations being genotyped as WT (see Supplementary Note). For each of the 1020 assayed promoter mutations, the corresponding fitness effect was estimated as the median of the selection coefficients observed across the four remaining replicates, which were considered unbiased.

The statistical significance of selection coefficients was assessed from the same four replicate measurements for each mutation, by comparing their mean to 0 using the SciPy (v 1.15.2)[105] implementation of the one-sample t-test, followed by Benjamini-Hochberg correction controlling the false discovery rate at 5%. For the coding sequence mutations added as controls, fitness effects were instead estimated as the median of the eight replicate measurements, without ignoring any of the corresponding samples. Alternatively, significant selection coefficients were also defined according to thresholds of minimal absolute deviations from 0, using the most extreme positive and negative replicates for each genotype. In this case, the number of fitness changes identified was compared with a null expectation derived from the observed experimental noise. To this end, the complete dataset of final selection coefficients (from the four culture replicates mentioned above) was shuffled into 1020 random "genotypes". From each culture replicate, 1020 selection coefficients were first sampled with replacement. The resulting four sets of values were then combined randomly into the 1020 genotypes, such that each was associated with four replicate measurements. The number of random genotypes exceeding the chosen threshold of absolute deviation from 0 was then compiled. This was repeated 10 000 times.

Comparison with published effects of amino acid substitutions in Fcy1p

The selection coefficients of added coding sequence mutations were first aggregated into grand medians, according to the corresponding amino acid substitutions. They were then compared with previously published variant effect scores obtained in the same conditions (12 µM 5-FC)[31]. The agreement between the two experiments was further assessed by classifying variants as resistant and susceptible in both cases. To this end, Gaussian mixture models with two components and a "tied" covariance matrix were fitted on each of the two datasets, using the scikit-learn (v 1.6.1)[106] implementation.

For the comparison of promoter and coding sequence mutations, the full dataset from[31] was used. All single-nucleotide substitutions in the reference sequence of the *FCY1* gene, obtained from the Saccharomyces Genome Database[98], were first performed computationally. The corresponding variant effects were retrieved from the Després *et al.* (2022) data[31], according to the resulting amino acid substitutions. Of 1431 possible single-nucleotide substitutions in *FCY1*, only 20 were not covered, mostly in the first and last codons. The Gaussian mixture model previously fitted on the full set of amino acid changes was then used to classify all single mutants as either resistant or susceptible. From this classification, the fraction of adaptive (resistance-conferring) mutations in the coding sequence was obtained.

Calculation of sort-seq expression scores

Within each of the six replicate sort-seq experiments, all variants which reached a total abundance of at least 85 reads across the five sampled bins were kept for downstream analysis. Although this included a small number of secondary mutations (outside of the F3-F4 mutagenized region), any variant which occurred in the mEGFP sequence was filtered out. For each of the remaining genotypes i, which includes the WT, fluorescence intensity was quantified from a weighted average B_i of bins where the corresponding reads were detected:

$$B_i = \frac{\sum_{j=0}^{4} j \frac{r_{i,j}}{r_{tot,j}} \frac{n_j}{n_{tot}}}{\sum_{j=0}^{4} \frac{r_{i,j}}{r_{tot,j}} \frac{n_j}{n_{tot}}}$$

In the above formula, j is the bin number, where 0 and 4 are associated with the lowest and highest fluorescence intensities, respectively. Variables $r_{i,j}$ and $r_{tot,j}$ are the read count for genotype i, as well as the total number of reads, in bin j. Finally, n_j and n_{tot} are respectively the number of cells collected from bin j and the total amount of cells collected across the five bins. The n_j/n_{tot} ratio was introduced because the number of yeast cells collected was not uniform across all bins.

Separately for each replicate experiment, the difference was computed between the B_m weighted average of each mutant m and the B_{WT} weighted average obtained for the WT genotype:

$$\Delta_{Ref,m} = B_m - B_{WT}$$

For each mutant, up to six replicate Δ_{Ref} scores were obtained (one for each sort-seq experiment). The median of these values was used as an estimate of each mutation's effect on fluorescence intensity. Positive values denote an increase relative to the WT, while negative values instead indicate a decrease.

Selection and construction of validation mutants

The sort-seq data was used to select 30 mutant genotypes for validation. These mutations were chosen to cover the whole range of scores, as well as all regions of the promoter where single-nucleotide changes appeared impactful. Substitutions were favored to ensure a high success rate to the downstream reconstruction of genotypes. To ensure that low-confidence mutations were not selected, a one-sample t-test (with Benjamini-Hochberg correction at FDR=5%) was first performed on the replicate sort-seq scores to identify Δ_{Ref} values which were significantly different from 0. This was also repeated with Δ_{mod} scores computed from B_{mod} weighted averages obtained from reads of bin 1 to 4, excluding the 0th one displaying the lowest fluorescence intensities. Because the Δ_{Ref} scores appeared biased towards positive values compared with Δ_{mod}, the latter were used to select significant mutations to be reconstructed, except when annotated *. This ensured a more stringent selection of genotypes displaying altered fluorescence intensity compared with the WT reference. Within each of the mutant groups described below, the specified number of sequences were chosen randomly, unless stated otherwise. Among mutations which significantly decreased Δ_{Mod} below zero, there were seven groups: 1) indels before the second methionine of *FCY1* (n=2); 2) indels after the second methionine of *FCY1* (n=2)*; 3) top 3 lowest scores between positions -50 and 0 of the promoter*; 4) substitutions between positions -133 and -127 (n=6); 5) substitutions between position -45 and -11 (n=2); 6) substitutions and insertions near the -75 position (one each) and 7) two chosen substitutions outside of regions with many significant effects. Among mutations which significantly increased Δ_{Mod} above zero, a similar approach was employed. The four categories were: 1) substitutions between position -125 and -118 (n=4); 2) substitutions between positions -112 to -109 (n=2); 3) substitutions between positions -63 to -57 (n=2) and 4) variants among the 25 highest scores (n=4).

The corresponding validation mutants were individually reconstructed using the CRISPR-Cas9 yeast transformation protocol described above. For each genotype, this was performed in two genetic backgrounds: fluorescent strain SA0247 (ΔF3-F4 with 42bp-mEGFP construct) and *FCY1* deletion (ΔF3-F4) mutant SA0159 (see **Supplementary Table 1**). A fusion PCR product harboring the single mutation of interest was used as donor DNA. This amplicon was generated by two consecutive rounds of PCR. In the first round, two reactions were performed

with WT genomic DNA as template. The first one used a generic forward primer hybridizing upstream of F3 (CLOP326-A11) with a mutant-specific reverse primer harboring the mutation of interest (**Supplementary Table 2**). In the second reaction, a forward primer with the mutation of interest (**Supplementary Table 2**) was used alongside a generic reverse primer binding either in the mEGFP sequence (CLOP326-G10; for background SA0247) or in the *FCY1* coding sequence (CLOP326-E10; for background SA0159). The two mutant-specific mutagenesis primers are complementary to each other, such that there is homology between the two amplicons. The second PCR round used the two PCR products from the first round diluted 1/40 as template and the same two generic primers used in the first pair of PCR (CLOP326-A11 with either CLOP326-G10 or CLOP326-E10, depending on the mutagenized background), which allowed fusion of both PCR parts. This approach thus allows to amplify the full F3-F4 promoter region (with a short segment of *FCY1*) while introducing a single mutation within it and adding homology regions on both ends for subsequent genomic insertion. It was additionally used to perform deletion of short segments of the promoter (-133 to -127 and -125 to -118), similarly in both genetic backgrounds. After transformation of each fusion PCR product, clones from each transformation were sent for Sanger sequencing to validate proper genomic integration of the mutation. All 66 resulting strains, including reconstructions of the WT promoter, are listed in **Supplementary Table 1**.

Expression measurements by flow cytometry

Measurement of promoter activity using fluorescent strains was performed with a Guava EasyCyte HT cytometer (CytekBio). Each strain of interest was grown overnight in three replicates in SC complete media at 30°C with shaking, along with similar replicate cultures of a reference strain with the WT promoter (SA0247 or AKD1462) and of a non-fluorescent control (AKD1123 or AKD1429). The overnight cultures were diluted at 0.15 OD_{595} in fresh SC-Trp (MSG) media and grown at 30°C with shaking until ~0.5 OD_{595}. Exponentially growing cells were then diluted to 0.05 OD_{595} in fresh media and single-cell fluorescence intensity measurements were performed with the Guava EasyCyte HT system. For each replicate culture of each strain, 5000 events were acquired (488 nm excitation; 525 nm detection). The medians of the corresponding distributions of fluorescence intensities were used to estimate relative expression levels.

Within each cytometry experiment, events were first filtered out according to forward scatter and side scatter values, with the 99th percentile of each distribution used as thresholds. All fluorescence intensities were then normalized by the corresponding forward scatter measurement. From these normalized fluorescence intensities, a median was computed for each replicate culture assayed. These medians were then normalized by the grand median of measurements made for the reference strain, such that they were converted into WT-relative fluorescence intensities. The resulting values, and especially their medians, were used as estimates of relative expression levels. Accordingly, three replicate measurements of WT-relative fluorescence intensity were obtained for each tested strain, with a value of 1 being equal to the WT level.

Calibration of sort-seq scores and bootstrapping of relative expression levels

WT-relative fluorescence intensities were first measured for all validation mutants reconstructed in the SA0247 (42bp-mEGFP) background, as described above. These

measurements were then compared with the corresponding sort-seq expression scores and a LOWESS regression was fitted. This approach allowed to simultaneously validate the sort-seq experiments and convert the resulting scores into more interpretable WT-relative fluorescence intensities. The Δ_{Ref} scores were used, since some genotypes with very low fluorescence intensities were absent from the Δ_{mod} dataset. In addition, because fluorescence intensity measurements from reconstructed mutants were used to fit the relationship, any bias in Δ_{Ref} was unlikely to affect the downstream inference of relative expression levels. So that uncertainties could also be obtained for each of the >1000 mutant expression levels, both the LOWESS fitting and the conversion of sort-seq scores into relative fluorescence intensities were performed within a bootstrapping routine.

At the start of each iteration, Δ_{Ref} scores were first resampled with replacement. For each genotype which had been included in the sort-seq analysis, as many values were sampled as there were replicate measurements among the six experiments. Per-genotype medians were then computed on the newly sampled data. For each reconstructed mutant for which fluorescence had been measured directly, three median WT-relative fluorescence intensities were sampled with replacement among the corresponding replicate measurements. Grand medians were computed from each trio of observations, which were combined with the newly recomputed median Δ_{Ref} scores of the corresponding validation genotypes to obtain 30 pairs. From these data points, LOWESS regression was performed with median Δ_{Ref} as independent variable, using the “nonparametric.lowess” implementation from statsmodels (v 0.14.4)[104] with default parameters. This regression was then evaluated at each of the >1000 median Δ_{Ref} values obtained from the first resampling steps. This generated an estimate of WT-relative fluorescence intensity for each of the genotypes which had been assayed in the sort-seq screen. The regression was also evaluated at 100 equally spaced values from the minimal to the maximal median Δ_{Ref} observed experimentally. This process was repeated 5000 times. After all iterations had been completed, medians and corresponding 95% prediction intervals were computed from the distribution of WT-relative fluorescence intensities obtained for each mutant genotype. These medians were used as WT-relative expression levels associated with the corresponding F3-F4 mutations. The same calculations were performed for each of the 100 equally spaced values at which the LOWESS had been evaluated, thus allowing to plot the relationship between Δ_{Ref} and relative fluorescence intensity.

Experimental measurements of the expression-fitness function

Fitting of dose-response curves

The WT background AKD1123 was phenotyped in liquid media containing various concentrations of 5-FC and cytosine, according to the growth assays protocol described above. From two to three replicates were made for each concentration of 5-FC and cytosine. A five-parameter logistic model was then fitted, using the exponential growth rate measurements and the concentrations of 5-FC / cytosine as dependent and independent variables, respectively. The resulting curves were converted into relative inhibition or relative growth, using measured growth rates in absence of 5-FC or at the highest cytosine concentration as references.

1201 **Construction of strains with distinct *FCY1* expression levels**
1202 Six yeast promoters covering a wide range of expression levels were used to construct two
1203 sets of strains, respectively from *FCY1* background AKD1123 and from fluorescent
1204 background SA0228. The six promoter cassettes, each including a G418 resistance marker,
1205 were PCR-amplified from the corresponding plasmids[95] (**Supplementary Table 3**) using
1206 oligonucleotides CLOP338-G6 and CLOP338-H6 (**Supplementary Table 2**). The two primers
1207 provided homology for genomic insertion immediately upstream of *FCY1* (or of its first 42 bp
1208 for the SA0228 background), between the gene and its native promoter. Yeast transformations
1209 were performed in the two genetic backgrounds as previously described, using each of the
1210 corresponding PCR products. Transformants were selected on YPD+HYG+G418 media.
1211 Correct integration of the promoter cassette was validated by PCR in four clones from each
1212 transformation, using oligonucleotides CLOP310-D1 and CLOP291-A3 (**Supplementary**
1213 **Table 2**). Two to three positive clones were obtained in most cases and validated by Sanger
1214 sequencing, such that biological replicates could be used in downstream assays for the
1215 majority of genotypes. All 25 corresponding yeast strains are listed in **Supplementary Table**
1216 **1**.
1217
1218 **Measurements of growth rates and fluorescence intensities**
1219 The set of *FCY1* strains (background AKD1123) was used to measure growth rates in liquid
1220 media containing different concentrations of 5-FC or cytosine, following the growth assays
1221 protocol described above. The set of fluorescent strains (background SA0228) was instead
1222 used in flow cytometry experiments to measure expression levels (see Expression
1223 measurements by flow cytometry). In both cases, all available biological replicates were used
1224 for each genotype. When less than three strains were available, three replicate cultures were
1225 nonetheless inoculated.
1226
1227 The growth assays in 5-FC media were performed in more than one experiment. To combine
1228 measurements obtained for concentrations of 12 μM and 3 μM (**Extended Data Fig. 9**), a
1229 normalization step was first performed.Growth rates from the first experiment were adjusted
1230 to have the same grand medians as those of the second one, by the addition of the difference
1231 between the two experiments' grand medians. This was performed separately for each
1232 concentration and allowed to combine the two sets of growth rates, in spite of potential
1233 systematic differences between them – for instance if the true concentration of 5-FC in the
1234 media varied. The cytometry measurements, which were done in parallel with the growth
1235 assays, were also performed in two separate experiments. In this case, potential systematic
1236 differences were controlled for by computing WT-relative fluorescence intensities within each
1237 experiment and only combining measurements after this step.
1238
1239 **Fitting of expression-fitness functions**
1240 Fluorescence measurements associated with each promoter were combined with the
1241 corresponding growth rates. Measurements made from the AKD1123 WT strain were however
1242 not used, because this strain lacks the G418 resistance marker included in promoter cassettes
1243 and is thus not directly comparable. An exponential function was fitted for each condition,
1244 using WT-relative fluorescence as the independent variable and growth rate as the dependent
1245 variable. This could model the sharp decline observed in 5-FC as well as the steep increase
1246 seen in cytosine conditions. The fitting of all expression-fitness curves was performed within
1247 a bootstrapping routine. For each promoter, two WT-relative fluorescence intensity
1248 measurements were first sampled with replacement from each cytometry experiment. Then,

1249 the four resulting values were resampled with replacement. Corresponding growth rates were
1250 sampled similarly. If a given condition had been tested in two experiments, two replicate
1251 growth rates were sampled with replacement from each experiment for each promoter, and
1252 the four measurements were afterwards resampled with replacement. This approach was
1253 employed to avoid favoring one experiment over the other, since the number of replicates
1254 sometimes varied between them. When only one experiment had been performed, replicate
1255 growth rates from this assay were instead resampled with replacement. Once relative
1256 fluorescence intensities and growth rates had been sampled, values obtained for the same
1257 promoter were paired randomly into data points. The exponential function was then fitted using
1258 the "optimize.curve_fit" implementation of the SciPy module (v 1.15.2)[105] module and
1259 evaluated over 250 equally spaced values covering the full range of WT-relative fluorescence
1260 intensities. After 5000 iterations, medians and corresponding 95% prediction intervals were
1261 computed for each of these 250 fluorescence levels, thus generating the expression-fitness
1262 function. This was repeated for each of the 5-FC and cytosine concentrations tested.
1263
1264 **Inference of the expected distribution of effects on fitness for promoter mutations**
1265 Changes in *FCY1* expression were assumed to have the same effect on fitness as variations
1266 in 5-FC concentration. The dose-response curve (**Extended Data Fig. 2**) was converted such
1267 that concentrations were expressed as relative changes from 12 µM (used in the bulk
1268 competition experiment). From this new curve, corresponding exponential growth rates were
1269 inferred for the 1019 relative expression estimates obtained from the sort-seq experiment. All
1270 growth rates were then converted into selection coefficients, as described above.
1271

1272 Validation of *FCY1* promoter variants fitness effects

1273 The previously described validation mutants reconstructed in the SA0159 *FCY1* background
1274 (**Supplementary Table 1**) were grown in SC complete (MSG) media with 12 µM 5-FC and
1275 without, according to the growth assays protocol described above. For all other mutants,
1276 growth rates measured in 12 µM 5-FC were converted into selection coefficients as described
1277 in the growth assays section, using growth rates measured in the same experiment for
1278 CRISPR reconstructions of the WT genotype in the SA0159 background (generated at the
1279 same time as the validation mutants). The expression-fitness function obtained in the same
1280 conditions and the bounds of its 95% prediction interval were also converted into selection
1281 coefficients using the same approach. In this case, the WT growth rate interpolated from the
1282 curve was instead used as comparison.
1283

1284 Assessment of 5-FC resistance arising from mutant library construction

1285 To test whether mutations conferring 5-FC resistance were introduced during mutant library
1286 construction, the process described in *Construction of mutant libraries* was repeated, but using
1287 WT sequences of the F3 and F4 regions. Briefly, the same two oligonucleotide pairs
1288 (CLOP299-C3/CLOP299-C4 and CLOP299-D3/CLOP299-D4) were first used to amplify F3
1289 and F4 from genomic DNA of the AKD1123 strain, using KAPA HiFi HotStart DNA polymerase
1290 (Roche Molecular Systems). Competent cells from strains SA0129 and SA0130 (prepared as
1291 described by [96]) were then transformed with the F3 and F4 inserts, respectively. Two biological
1292 replicates of each transformation were made. This was performed following the CRISPR-Cas9
1293 approach described above for the pools of single mutations. After 48 hours of incubation at
1294 30°C, transformants from each Petri dish were collected in 5 ml YPD by scraping, as

previously described. From each of the resulting cell suspension, cultures were started in 5 ml of SC complete [pH=6] + HYG, from 100 µl of cells.

After overnight incubation at 30°C with shaking, OD_{595} was measured for each culture and dilutions at $OD_{595} = 1$ were made in 200 µl of sterile water. Additional dilutions were afterwards made for OD_{595} values from 0.2 to 0.002. From each dilution, 150 µl were plated on SC complete (MSG) solid media with and without 5-FC [100 µg/ml]. Petri dishes were incubated at 30°C for 48 hours and colonies were counted. While colonies were too numerous to be counted on SC complete controls, some dilutions plated on 5-FC media resulted in countable colony numbers (between 20 and 340). For each plating (one per CRISPR-Cas9 transformation), colony counting was performed for two dilution levels, which acted as technical replicates. Since the SA0130 background was initially resistant to 5-FC due to the insertion of a natNT2 in place of F4, all associated platings were replica plated on YPD+NAT media. Nourseothricin-resistant colonies were subtracted from the total number of 5-FC-resistant colonies. The frequency of 5-FC resistance was then estimated as follows, separately for each plating on media containing 5-FC:

$$F_{LOF} = \frac{n_{FC}}{OD_d \, V_{plate} \, c_{ml}}$$

In the equation above, n_{FC} is the number of colonies on 5-FC media (after subtracting NAT resistant ones, when applicable), while OD_d is the OD_{595} of the plated dilution, V_{plate} is the volume deposited on each plate (0.15 ml) and c_{ml} is the number of yeast cells in 1 ml of media at $OD_{595} = 1$, for which a value of 10 million cells/ml was used. The resulting F_{LOF} was considered as a putative frequency of *FCY1* LOF in the corresponding transformations. To validate whether that was the case, 16 colonies were selected for Sanger sequencing, using primer CLOP326-F10, which covered nearly 70% of the *FCY1* coding sequence **(Supplementary Table 2)**. All colony counts, resistance rates and relevant calculations are listed in **Supplementary Table 7**.

Simulations with *invisible* resistance-conferring mutations

A simulation framework replicating the bulk competition experiment was designed. It involved a library of neutral variants within which a small number of invisible resistance-conferring mutations was introduced, thus mimicking secondary mutations occurring outside of the genotyping region in the experiment. A set of variants of interest (hereafter named genotype labels) was first generated by sampling 1000 neutral mutations from $s \sim \mathcal{N}(0, 0.004)$ four times. The standard deviation used was set according to what was observed in sequencing libraries from cultures without 5-FC. Four transformations were then obtained by replicate random samplings of 10 000 clones, among which 17% were defined as WT ($s = 0$) and the remaining were drawn from the set of genotype labels. Within each such transformation replicate, a small fraction of clones was defined as harboring an invisible resistance-conferring mutation (or resistance flag), which increased its s (between 0.2 to 0.4, depending on simulations) without changing its genotype label. From each transformation, two replicate cultures were generated by randomly sampling two sets of 2.5 million cells, resulting in eight cultures, as in the bulk competition experiment (5-FC condition without added coding sequence variants). From each culture, 1 million of sequencing reads (genotype labels) were sampled randomly, to simulate sequencing of the initial T0 samples. Simulated cells in each culture were next aggregated according to genotype labels and presence/absence of a resistance flag, prior to calculation of WT-relative frequencies for each group (called true

genotypes). Growth over the course of a bulk competition experiment was simulated by modeling exponential changes of WT-relative frequencies through time for each true genotype i, as follows:

$$F_{i,T+1} = F_{i,T}\, e^{s_i n_{gen}}$$

The WT-relative frequency F of true genotype i at timepoint $T+1$ is obtained from the corresponding WT-relative frequency at previous timepoint T according to the genotype's true selection coefficient (increased by a resistance-conferring mutation when applicable) and the time interval n_{gen} between the two timepoints, in WT generations. Three such timepoints, from T1 to T3, were simulated and their respective durations were set in accordance with the bulk competition experiment. At each timepoint, 1 million reads as well as a new initial population of 2.5 million cells were sampled from each culture, as described above. Selection coefficients for all variants were then computed from time-series of simulated read counts, as described for experimental variant frequencies. This was performed from genotype labels only as well as with knowledge of true genotypes (resistance flags). Such simulations of one full experiment were repeated while varying both the initial frequency and the fitness effect of resistance-conferring mutations. For each combination, multiple replicate simulated experiments (each with eight simulated cultures) were performed.

Data availability

All sequencing data from bulk competition and sort-seq experiments has been deposited on the NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/PRJNA1321317). All other experimental data and files needed to replicate the analysis have been deposited on Zenodo: https://doi.org/10.5281/zenodo.17088129.

Code availability

All analyses were performed using custom Python scripts and notebooks, which are available at: https://github.com/Landrylab/Aube_et_al_2025_FCY1promoter. A custom Python pipeline was deployed on a computing cluster (Institut de Biologie Intégrative et des Systèmes, Université Laval) to compute variant counts and selection coefficients from raw sequencing data. As described above, this used the following software: FastQC (v 0.12.1), MultiQC (v 1.12)[100], PANDAseq (v 2.11)[101], VSEARCH (v 2.28.1)[102] and EMBOSS suite (v 6.5.7)[103]. The versions of all Python modules used are included as requirements files, separately for the analysis pipeline and for the notebooks, which were executed locally.

Acknowledgements

We thank Philippe Després, Isabelle Hatin, Olivier Namy and all Landry lab members, especially Soham Dibyachintan, David Jordan, Pascale Lemieux and Camille Bédard, for helpful discussions. We are grateful to Pavithra Venkataraman, Angel Cisneros, Soham Dibyachintan and Philippe Després, who provided comments on this manuscript. We additionally thank Angel Cisneros for helpful suggestions on the simulation framework. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN-2020-04844) for CRL. CRL holds the Canada Research Chair in Cellular Systems and Synthetic Biology. SA

was supported by doctoral research fellowships from Natural Sciences and Engineering Research Council of Canada and Fonds de recherche du Québec - Nature et technologies. The F65V-34a-mEGFP plasmid used in this study was a kind gift from Dr Michael Springer.

Author contributions

C.R.L., A.K.D and S.A. designed the study. S.A. and A.K.D. performed the experiments as well as the data analysis. S.A., A.K.D. and C.R.L. interpreted the results. S.A., A.K.D and C.R.L. wrote the manuscript.

References

1. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282–1294.e18 (2016).
2. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203–215 (2021).
3. McManus, K. F. *et al.* Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genet.* **13**, e1006560 (2017).
4. Tournamille, C., Colin, Y., Cartron, J. P. & Le Van Kim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat. Genet.* **10**, 224–228 (1995).
5. Breeyear, J. H. *et al.* Adaptive selection at G6PD and disparities in diabetes complications. *Nat. Med.* **30**, 2480–2488 (2024).
6. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–305 (2010).
7. Bédard, C. *et al.* FungAMR: a comprehensive database for investigating fungal mutations associated with antimicrobial resistance. *Nat. Microbiol.* (2025) doi:10.1038/s41564-025-02084-7.
8. Guerrero-Preston, R. *et al.* Key tumor suppressor genes inactivated by ‘greater promoter’ methylation and somatic mutations in head and neck cancer. *Epigenetics* **9**, 1031–1046 (2014).

1416 9. Behe, M. J. Experimental evolution, loss-of-function mutations, and 'the first rule of
1417 adaptive evolution'. *Q. Rev. Biol.* **85**, 419–445 (2010).
1418 10. Gerstein, A. C., Lo, D. S. & Otto, S. P. Parallel Genetic Changes and Nonparallel Gene–
1419 Environment Interactions Characterize the Evolution of Drug Resistance in Yeast.
1420 *Genetics* **192**, 241–252 (2012).
1421 11. Kvitek, D. J. & Sherlock, G. Whole genome, whole population sequencing reveals that
1422 loss of signaling networks is the major adaptive strategy in a constant environment.
1423 *PLoS Genet.* **9**, e1003972 (2013).
1424 12. Lang, G. I. & Desai, M. M. The spectrum of adaptive mutations in experimental
1425 evolution. *Genomics* **104**, 412–416 (2014).
1426 13. Lind, P. A., Libby, E., Herzog, J. & Rainey, P. B. Predicting mutational routes to new
1427 adaptive phenotypes. *Elife* **8**, e38822 (2019).
1428 14. Murray, A. W. Can gene-inactivating mutations lead to evolutionary novelty? *Curr. Biol.*
1429 **30**, R465–R471 (2020).
1430 15. Durand, R., Jalbert-Ross, J., Fijarczyk, A., Dubé, A. K. & Landry, C. R. Cross-feeding
1431 affects the target of resistance evolution to an antifungal drug. *PLoS Genet.* **19**,
1432 e1011002 (2023).
1433 16. Chen, P. & Zhang, J. The loci of environmental adaptation in a model eukaryote. *Nat.*
1434 *Commun.* **15**, 5672 (2024).
1435 17. Hodgins-Davis, A., Duveau, F., Walker, E. A. & Wittkopp, P. J. Empirical measures of
1436 mutational effects define neutral models of regulatory evolution in Saccharomyces
1437 cerevisiae. *Proceedings of the National Academy of Sciences* **116**, 21085–21093
1438 (2019).
1439 18. Vaishnav, E. D. *et al.* The evolution, evolvability and engineering of gene regulatory
1440 DNA. *Nature* **603**, 455–463 (2022).
1441 19. Tsuru, S., Hatanaka, N. & Furusawa, C. Promoters constrain evolution of expression
1442 levels of essential genes in Escherichia coli. *Mol. Biol. Evol.* **41**, msae185 (2024).
1443 20. Patwardhan, R. P. *et al.* High-resolution analysis of DNA regulatory elements by

1444 synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
1445 21. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian
1446 enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
1447 22. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in
1448 human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277
1449 (2012).
1450 23. Kircher, M. *et al.* Saturation mutagenesis of twenty disease-associated regulatory
1451 elements at single base-pair resolution. *Nat. Commun.* **10**, 3583 (2019).
1452 24. Metzger, B. P. H. *et al.* Contrasting Frequencies and Effects of cis- and trans-
1453 Regulatory Mutations Affecting Gene Expression. *Mol. Biol. Evol.* **33**, 1131–1146
1454 (2016).
1455 25. Duveau, F., Toubiana, W. & Wittkopp, P. J. Fitness effects of Cis-regulatory variants in
1456 the Saccharomyces cerevisiae TDH3 promoter. *Mol. Biol. Evol.* **34**, 2908–2912 (2017).
1457 26. Siddiq, M. A., Duveau, F. & Wittkopp, P. J. Plasticity and environment-specific
1458 relationships between gene expression and fitness in Saccharomyces cerevisiae. *Nat.*
1459 *Ecol. Evol.* **8**, 2184–2194 (2024).
1460 27. Rich, M. S. *et al.* Comprehensive Analysis of the SUL1 Promoter of Saccharomyces
1461 cerevisiae. *Genetics* **203**, 191–202 (2016).
1462 28. Ringel, A. E. *et al.* Yeast Tdh3 (glyceraldehyde 3-phosphate dehydrogenase) is a Sir2-
1463 interacting factor that regulates transcriptional silencing and rDNA recombination. *PLoS*
1464 *Genet.* **9**, e1003871 (2013).
1465 29. Kankipati, H. N., Rubio-Texeira, M., Castermans, D., Diallinas, G. & Thevelein, J. M.
1466 Sul1 and Sul2 sulfate transceptors signal to protein kinase A upon exit of sulfur
1467 starvation. *J. Biol. Chem.* **290**, 10430–10446 (2015).
1468 30. Vermes, A., Guchelaar, H. J. & Dankert, J. Flucytosine: a review of its pharmacology,
1469 clinical indications, pharmacokinetics, toxicity and drug interactions. *J. Antimicrob.*
1470 *Chemother.* **46**, 171–179 (2000).
1471 31. Després, P. C. *et al.* Asymmetrical dose responses shape the evolutionary trade-off

1472 between antifungal resistance and nutrient use. *Nat. Ecol. Evol.* **6**, 1501–1515 (2022).
1473 32. Lin, Z., Wu, W.-S., Liang, H., Woo, Y. & Li, W.-H. The spatial distribution of cis
1474 regulatory elements in yeast promoters and its implications for transcriptional regulation.
1475 *BMC Genomics* **11**, 581 (2010).
1476 33. Mahendrawada, L., Warfield, L., Donczew, R. & Hahn, S. Low overlap of transcription
1477 factor DNA binding and regulatory targets. *Nature* **642**, 796–804 (2025).
1478 34. Lemieux, P., Bradley, D., Dubé, A. K., Dionne, U. & Landry, C. R. Dissection of the role
1479 of a Src homology 3 domain in the evolution of binding preference of paralogous
1480 proteins. *Genetics* **226**, iyad175 (2024).
1481 35. Dionne, U. *et al.* Protein context shapes the specificity of SH3 domain-mediated
1482 interactions in vivo. *Nat. Commun.* **12**, 1597 (2021).
1483 36. Matuszewski, S., Hildebrandt, M. E., Ghenu, A.-H., Jensen, J. D. & Bank, C. A statistical
1484 guide to the design of deep mutational scanning experiments. *Genetics* **204**, 77–87
1485 (2016).
1486 37. de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million
1487 random promoters. *Nat. Biotechnol.* **38**, 56–65 (2020).
1488 38. Peterman, N. & Levine, E. Sort-seq under the hood: implications of design choices on
1489 large-scale characterization of sequence-function relations. *BMC Genomics* **17**, 206
1490 (2016).
1491 39. de Boer, C. G. & Hughes, T. R. YeTFaSCo: a database of evaluated yeast transcription
1492 factor sequence specificities. *Nucleic Acids Res.* **40**, D169–79 (2012).
1493 40. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-
1494 initiation complexes. *Nature* **483**, 295–301 (2012).
1495 41. McMillan, J., Lu, Z., Rodriguez, J. S., Ahn, T.-H. & Lin, Z. YeasTSS: an integrative web
1496 database of yeast transcription start sites. *Database (Oxford)* **2019**, baz048 (2019).
1497 42. Eisenberg, A. R. *et al.* Translation initiation site profiling reveals widespread synthesis of
1498 non-AUG-initiated protein isoforms in yeast. *Cell Syst.* **11**, 145–160.e5 (2020).
1499 43. Huang, Q., Szklarczyk, D., Wang, M., Simonovic, M. & von Mering, C. PaxDb 5.0:

1500 Curated protein quantification data suggests adaptive proteome changes in yeasts. *Mol.*
1501 *Cell. Proteomics* **22**, 100640 (2023).
1502 44. Erb, I. & van Nimwegen, E. Transcription factor binding site positioning in yeast:
1503 proximal promoter motifs characterize TATA-less promoters. *PLoS One* **6**, e24279
1504 (2011).
1505 45. Park, D., Morris, A. R., Battenhouse, A. & Iyer, V. R. Simultaneous mapping of transcript
1506 ends at single-nucleotide resolution and identification of widespread promoter-
1507 associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* **42**, 3736–
1508 3749 (2014).
1509 46. Pelechano, V., Wei, W. & Steinmetz, L. M. Extensive transcriptional heterogeneity
1510 revealed by isoform profiling. *Nature* **497**, 127–131 (2013).
1511 47. Struhl, K. Naturally occurring poly(dA-dT) sequences are upstream promoter elements
1512 for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8419–8423
1513 (1985).
1514 48. de Boer, C. G. & Hughes, T. R. Poly-dA:dT tracts form an in vivo nucleosomal turnstile.
1515 *PLoS One* **9**, e110479 (2014).
1516 49. Ichikawa, Y. *et al.* Sequence-directed nucleosome-depletion is sufficient to activate
1517 transcription from a yeast core promoter in vivo. *Biochem. Biophys. Res. Commun.* **476**,
1518 57–62 (2016).
1519 50. Jansen, A., van der Zande, E., Meert, W., Fink, G. R. & Verstrepen, K. J. Distal
1520 chromatin structure influences local nucleosome positions and gene expression. *Nucleic*
1521 *Acids Res.* **40**, 3870–3885 (2012).
1522 51. Delma, F. Z. *et al.* Genetic mutations in Cryptococcus neoformans pyrimidine salvage
1523 pathway enzymes contribute to reduced susceptibility against 5-fluorocytosine. *NPJ*
1524 *Antimicrob. Resist.* **2**, 24 (2024).
1525 52. Kruglyak, L. *et al.* Insufficient evidence for non-neutrality of synonymous mutations.
1526 *Nature* **616**, E8–E9 (2023).
1527 53. Wyman, S. K. *et al.* Synthesis errors in oligonucleotides are propagated into the

genome by homology-directed repair. *bioRxiv* 2025.07.08.663753 (2025) doi:10.1101/2025.07.08.663753.

54. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).

55. Rossi, M. J. *et al.* A high-resolution protein architecture of the budding yeast genome. *Nature* **592**, 309–314 (2021).

56. Jin, Z. *et al.* Unraveling the regulatory dynamics of bidirectional promoters for modulating gene co-expression and metabolic flux in Saccharomyces cerevisiae. *Nucleic Acids Res.* **53**, gkaf511 (2025).

57. Zhang, C. *et al.* Yeast9: a consensus genome-scale metabolic model for S. cerevisiae curated by the community. *Mol. Syst. Biol.* **20**, 1134–1150 (2024).

58. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. A. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2310–8 (2014).

59. Kemble, H., Nghe, P. & Tenaillon, O. Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. *Evol. Appl.* **12**, 1721–1742 (2019).

60. Kacser, H. & Burns, J. A. The molecular basis of dominance. *Genetics* **97**, 639–666 (1981).

61. Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. Metabolic flux and fitness. *Genetics* **115**, 25–31 (1987).

62. Jiang, L., Mishra, P., Hietpas, R. T., Zeldovich, K. B. & Bolon, D. N. A. Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS Genet.* **9**, e1003600 (2013).

63. Rest, J. S. *et al.* Nonlinear fitness consequences of variation in expression level of a eukaryotic gene. *Mol. Biol. Evol.* **30**, 448–456 (2013).

64. Hartl, D. L., Dykhuizen, D. E. & Dean, A. M. Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655–674 (1985).

65. Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, aaf2786–aaf2786 (2016).

1556 66. Li, X., Walhout, A. J. M. & Yilmaz, L. S. Enhanced flux potential analysis links changes
1557 in enzyme expression to metabolic flux. *Mol. Syst. Biol.* **21**, 413–445 (2025).
1558 67. Cools, H. J., Hawkins, N. J. & Fraaije, B. A. Constraints on the evolution of azole
1559 resistance in plant pathogenic fungi. *Plant Pathol.* **62**, 36–42 (2013).
1560 68. Todd, R. T. & Selmecki, A. Expandable and reversible copy number amplification drives
1561 rapid adaptation to antifungal drugs. *Elife* **9**, e58349 (2020).
1562 69. Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to
1563 controlled nutrient-limited environments in yeast. *PLoS Genet.* **4**, e1000303 (2008).
1564 70. Payen, C. *et al.* The dynamics of diverse segmental amplifications in populations of
1565 Saccharomyces cerevisiae adapting to strong selection. *G3 (Bethesda)* **4**, 399–409
1566 (2014).
1567 71. Risch, T. *et al.* YgiV promoter mutations cause resistance to cystobactamids and
1568 reduced virulence factor expression in Escherichia coli. *NPJ Antimicrob. Resist.* **2**, 33
1569 (2024).
1570 72. Händel, N., Schuurmans, J. M., Feng, Y., Brul, S. & ter Kuile, B. H. Interaction between
1571 mutations and regulation of gene expression during development of de novo antibiotic
1572 resistance. *Antimicrob. Agents Chemother.* **58**, 4371–4379 (2014).
1573 73. Siu, L. K., Ho, P. L., Yuen, K. Y., Wong, S. S. & Chau, P. Y. Transferable
1574 hyperproduction of TEM-1 beta-lactamase in Shigella flexneri due to a point mutation in
1575 the pribnow box. *Antimicrob. Agents Chemother.* **41**, 468–470 (1997).
1576 74. Sharon, E. *et al.* Functional genetic variants revealed by massively parallel precise
1577 genome editing. *Cell* **175**, 544–557.e16 (2018).
1578 75. Vanella, R. *et al.* Understanding activity-stability tradeoffs in biocatalysts by enzyme
1579 proximity sequencing. *Nat. Commun.* **15**, 1807 (2024).
1580 76. Amorosi, C. J. *et al.* Massively parallel characterization of CYP2C9 variant enzyme
1581 activity and abundance. *Am. J. Hum. Genet.* **108**, 1735–1751 (2021).
1582 77. Prywes, N. *et al.* A map of the rubisco biochemical landscape. *Nature* **638**, 823–828
1583 (2025).

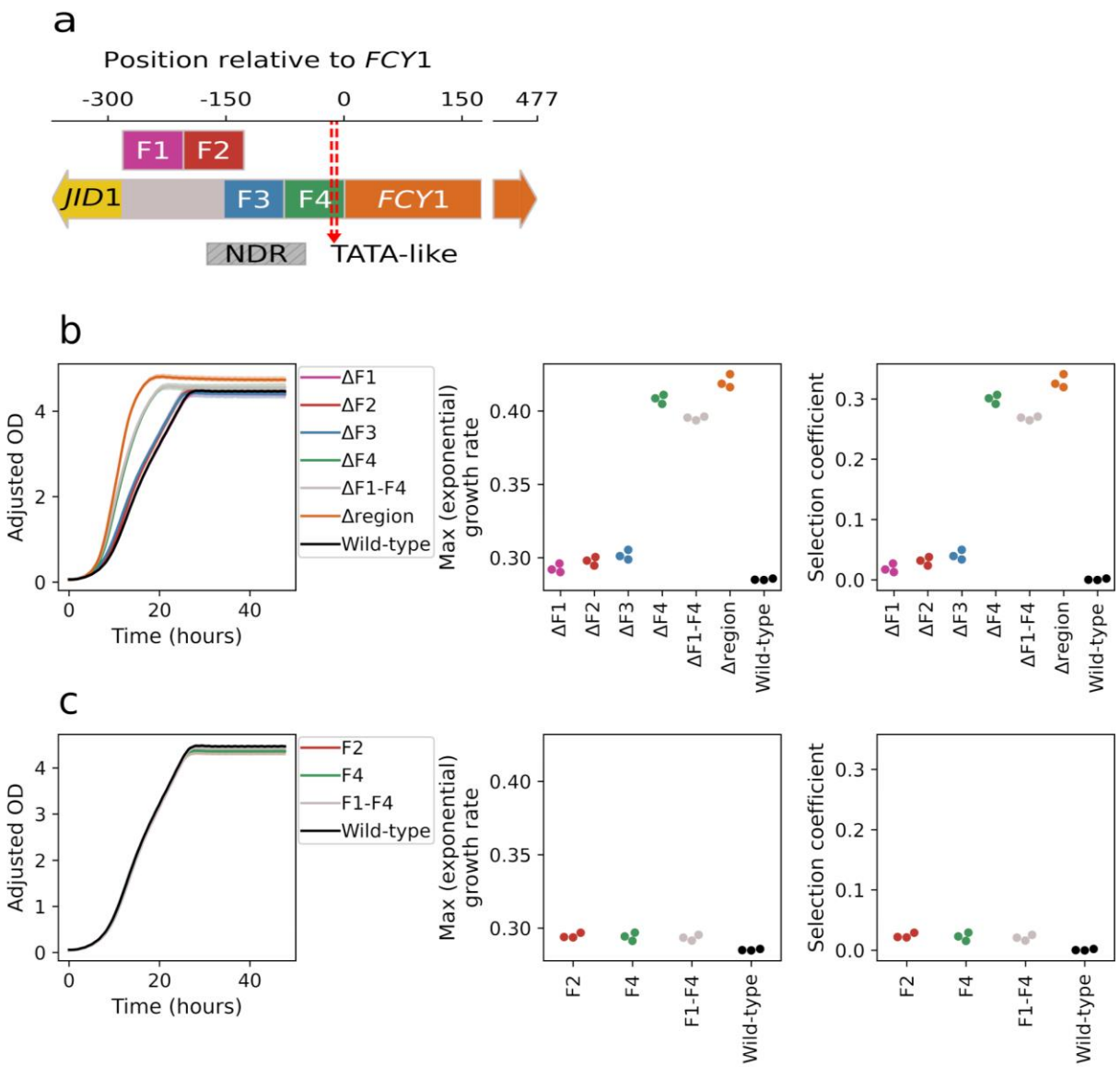
1584 78. Grønbæk-Thygesen, M. *et al.* Deep mutational scanning reveals a correlation between
1585 degradation and toxicity of thousands of aspartoacylase variants. *Nat. Commun.* **15**,
1586 4026 (2024).
1587 79. Beltran, A., Jiang, X. 'er, Shen, Y. & Lehner, B. Site-saturation mutagenesis of 500
1588 human protein domains. *Nature* **637**, 885–894 (2025).
1589 80. Bédard, C. *et al.* Most azole resistance mutations in the Candida albicans drug target
1590 confer cross-resistance without intrinsic fitness cost. *Nat. Microbiol.* **9**, 3025–3040
1591 (2024).
1592 81. Tabet, D. R. *et al.* The functional landscape of coding variation in the familial
1593 hypercholesterolemia gene LDLR. *Science* eady7186 (2025).
1594 82. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by
1595 RNA sequencing. *Science* **320**, 1344–1349 (2008).
1596 83. Luthra, I. *et al.* Regulatory activity is the default DNA state in eukaryotes. *Nat. Struct.*
1597 *Mol. Biol.* **31**, 559–567 (2024).
1598 84. Meneu, L. *et al.* Sequence-dependent activity and compartmentalization of foreign DNA
1599 in a eukaryotic nucleus. *Science* **387**, eadm9466 (2025).
1600 85. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of
1601 adaptation: The locus of evolution. *Evolution* **61**, 995–1016 (2007).
1602 86. Brown, K. M., Depristo, M. A., Weinreich, D. M. & Hartl, D. L. Temporal constraints on
1603 the incorporation of regulatory mutants in evolutionary pathways. *Mol Biol Evol* **26**,
1604 2455–2462 (2009).
1605 87. Otwinowski, J., McCandlish, D. M. & Plotkin, J. B. Inferring the shape of global epistasis.
1606 *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7550–E7558 (2018).
1607 88. Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature*
1608 **533**, 397–401 (2016).
1609 89. Wylie, C. S. & Shakhnovich, E. I. A biophysical protein folding model accounts for most
1610 mutational fitness effects in viruses. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9916–9921
1611 (2011).

1612 90. Serohijos, A. W. R. & Shakhnovich, E. I. Merging molecular mechanism and evolution:
1613 theory and computation at the interface of biophysics and evolutionary population
1614 genetics. *Curr. Opin. Struct. Biol.* **26**, 84–91 (2014).
1615 91. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr.*
1616 *Opin. Struct. Biol.* **19**, 596–604 (2009).
1617 92. Milind, N. *et al.* Buffering and non-monotonic behavior of gene dosage response curves
1618 for human complex traits. *medRxiv* 2024.11.11.24317065 (2025)
1619 doi:10.1101/2024.11.11.24317065.
1620 93. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces
1621 cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene
1622 disruption and other applications. *Yeast* **14**, 115–132 (1998).
1623 94. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS
1624 carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
1625 95. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new
1626 fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–
1627 962 (2004).
1628 96. Ryan, O. W., Poddar, S. & Cate, J. H. D. CRISPR-Cas9 Genome Engineering in
1629 Saccharomyces cerevisiae Cells. *Cold Spring Harb. Protoc.* **2016**, db.prot086827
1630 (2016).
1631 97. Chevin, L.-M. On measuring selection in experimental evolution. *Biol. Lett.* **7**, 210–213
1632 (2011).
1633 98. Engel, S. R. *et al.* Saccharomyces Genome Database: advances in genome annotation,
1634 expanded biochemical pathways, and other key enhancements. *Genetics* **229**, iyae185
1635 (2025).
1636 99. Amberg, D. C., Burke, D. J. & Strathern, J. N. *Methods in Yeast Genetics 2005: A Cold*
1637 *Spring Harbor Laboratory Course Manual*. (Cold Spring Harbor Laboratory Press, New
1638 York, NY, 2005).
1639 100. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results

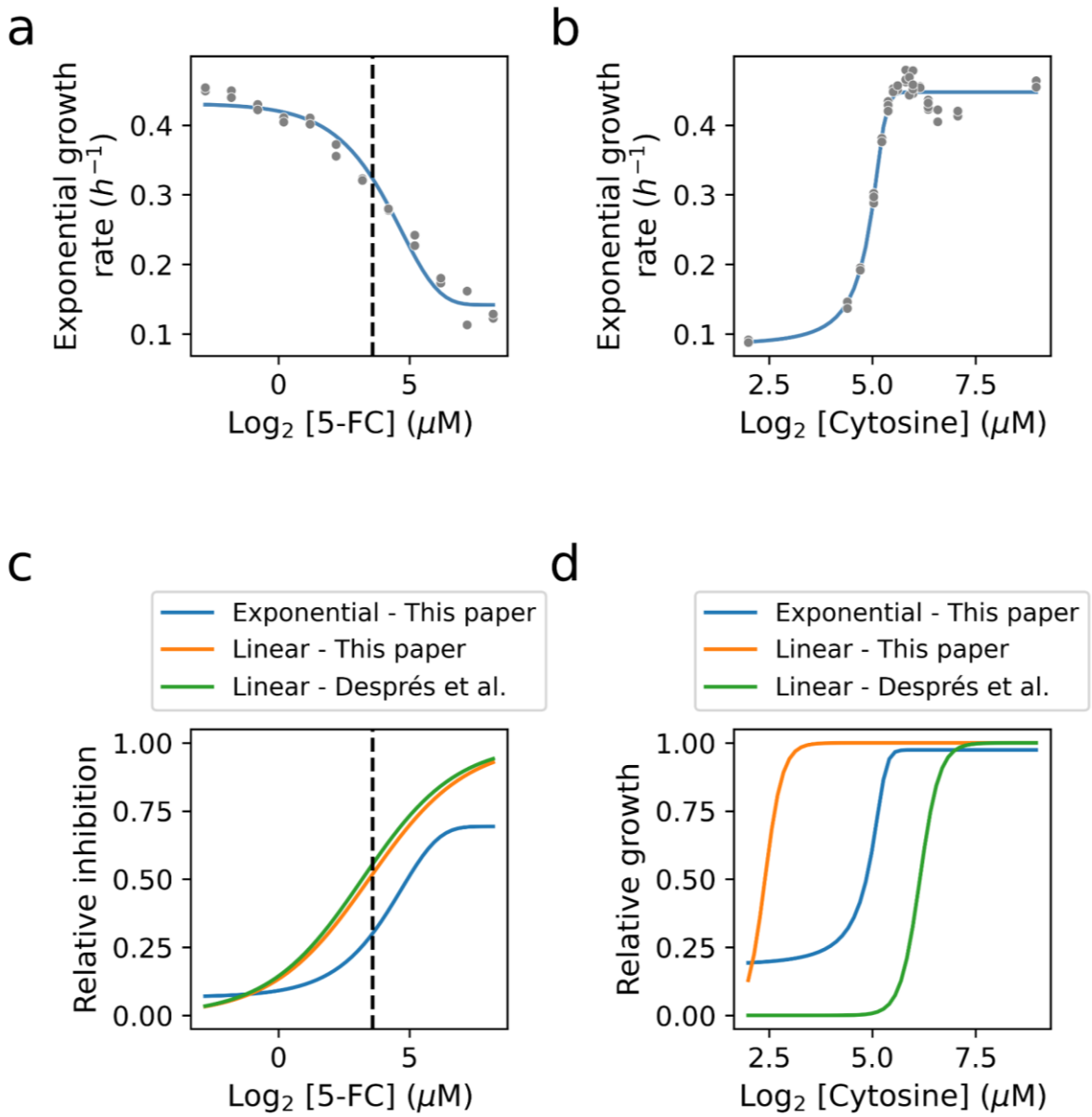
1640 for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
1641 101. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D.
1642 PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31
1643 (2012).
1644 102. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open
1645 source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
1646 103. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European molecular biology open
1647 software suite. *Trends Genet.* **16**, 276–277 (2000).
1648 104. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with
1649 python. in *9th Python in Science Conference* (2010).
1650 105. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python.
1651 *Nat. Methods* **17**, 261–272 (2020).
1652 106. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine*
1653 *Learning Research* **12**, 2825–2830 (2011).
1654 107. Chereji, R. V., Ramachandran, S., Bryson, T. D. & Henikoff, S. Precise genome-wide
1655 mapping of single nucleosomes and linkers in vivo. *Genome Biol.* **19**, 19 (2018).

1656
1657
1658
1659
1660
1661
1662
1663
1664

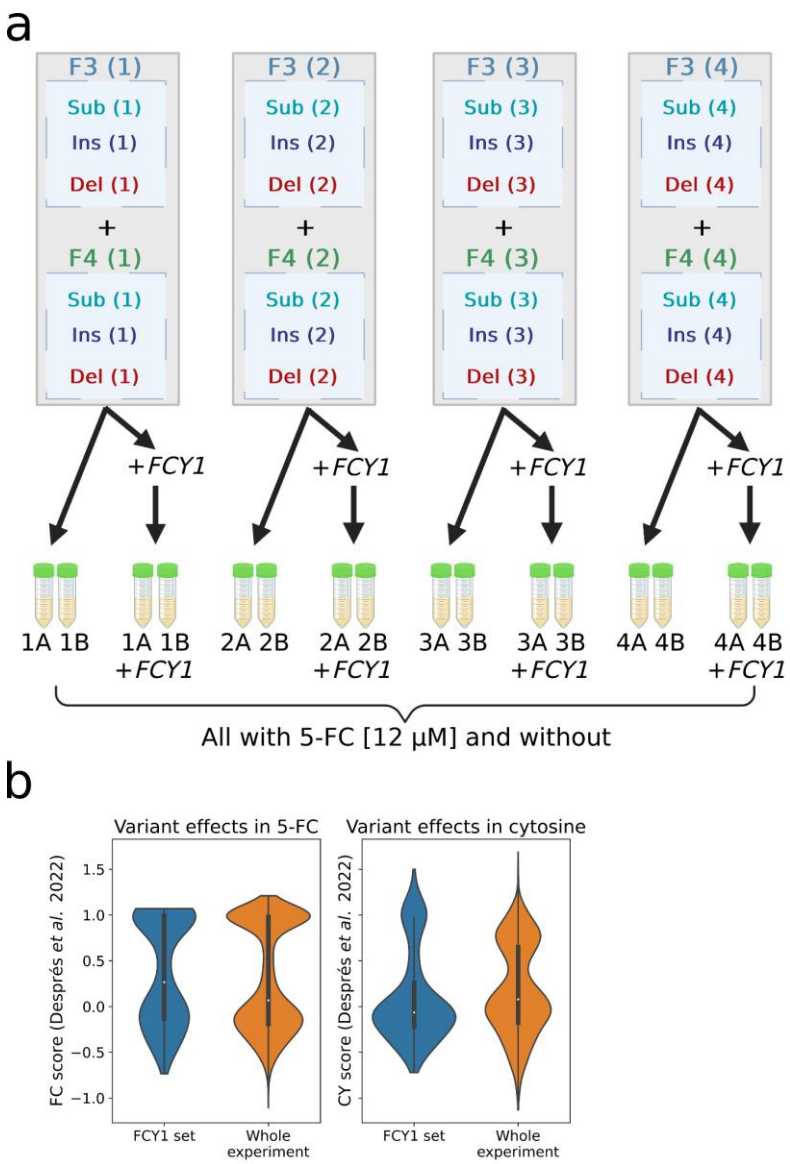
1665 **Extended Data Figures:**



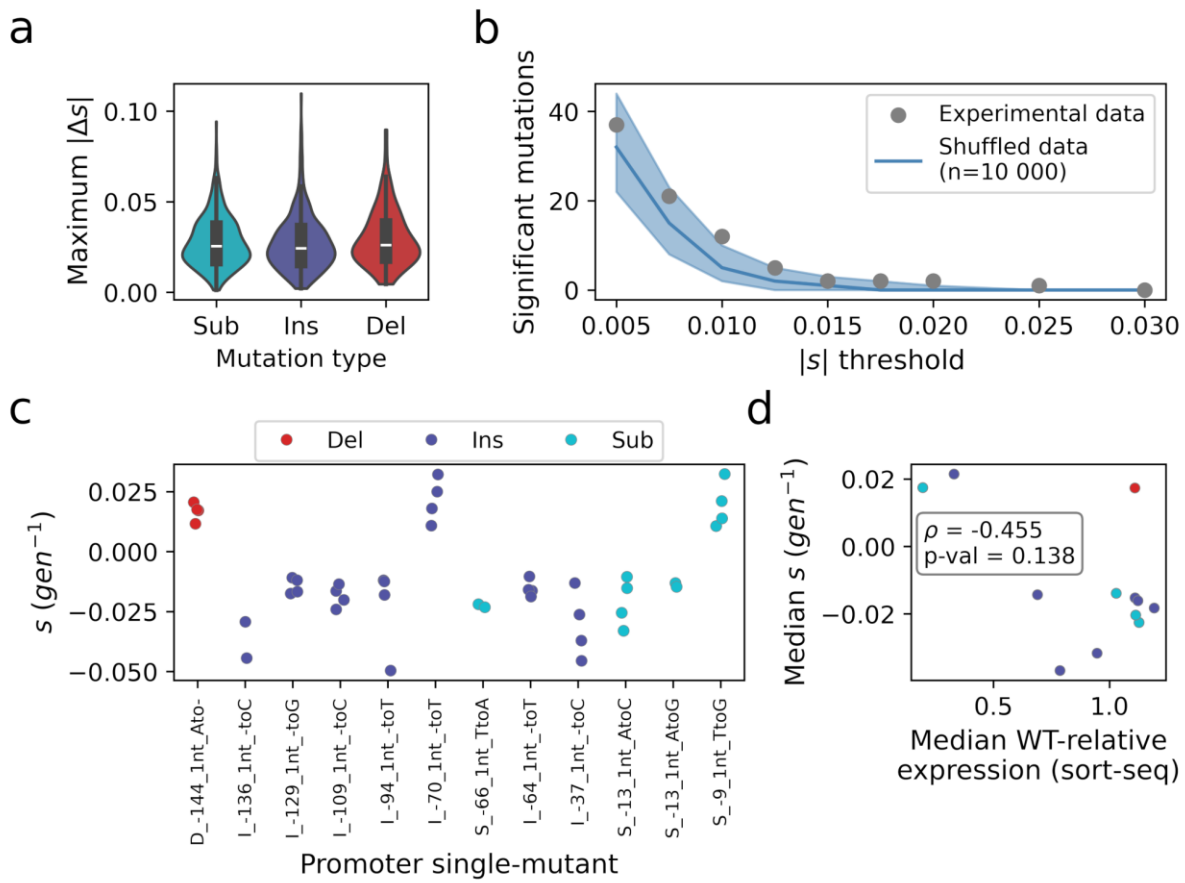
1666
1667 **Extended Data Fig. 1**: **The deletion of the F4 promoter fragment is clearly adaptive in 5-FC**
1668 **conditions. a,** Schematics of the four 77 bp fragments into which the intergenic region between *JID1*
1669 (YPR061C) and *FCY1* (YPR062W) was divided. There is a 26 bp overlap between F2 and F3. The
1670 nucleosome-depleted region (NDR) of the corresponding bidirectional promoter[56,107] as well as the
1671 TATA-like element, involved in transcription initiation at *FCY1*[40,41], are shown. **b,** Fitness of deletion
1672 mutants in [12 μM] 5-FC. The ΔF1-F4 mutant refers to deletion of the complete intergenic region, while
1673 Δregion instead represents the simultaneous deletion of the intergenic region and the *FCY1* gene.
1674 Exponential growth rates and the corresponding selection coefficients (*Methods*) show that F4 is the
1675 only promoter fragment for which deletion confers a clear growth advantage (and thus 5-FC resistance).
1676 **c,** Fitness of reconstructed WT strains in [12 μM] 5-FC. Here, F2, F4 and F1-F4 refer to CRISPR
1677 reconstruction of the WT intergenic region from the corresponding deletion mutant (*Methods*). In all
1678 three cases, the WT 5-FC sensitivity is recovered. All growth measurements were made from three
1679 replicate cultures of each strain.
1680



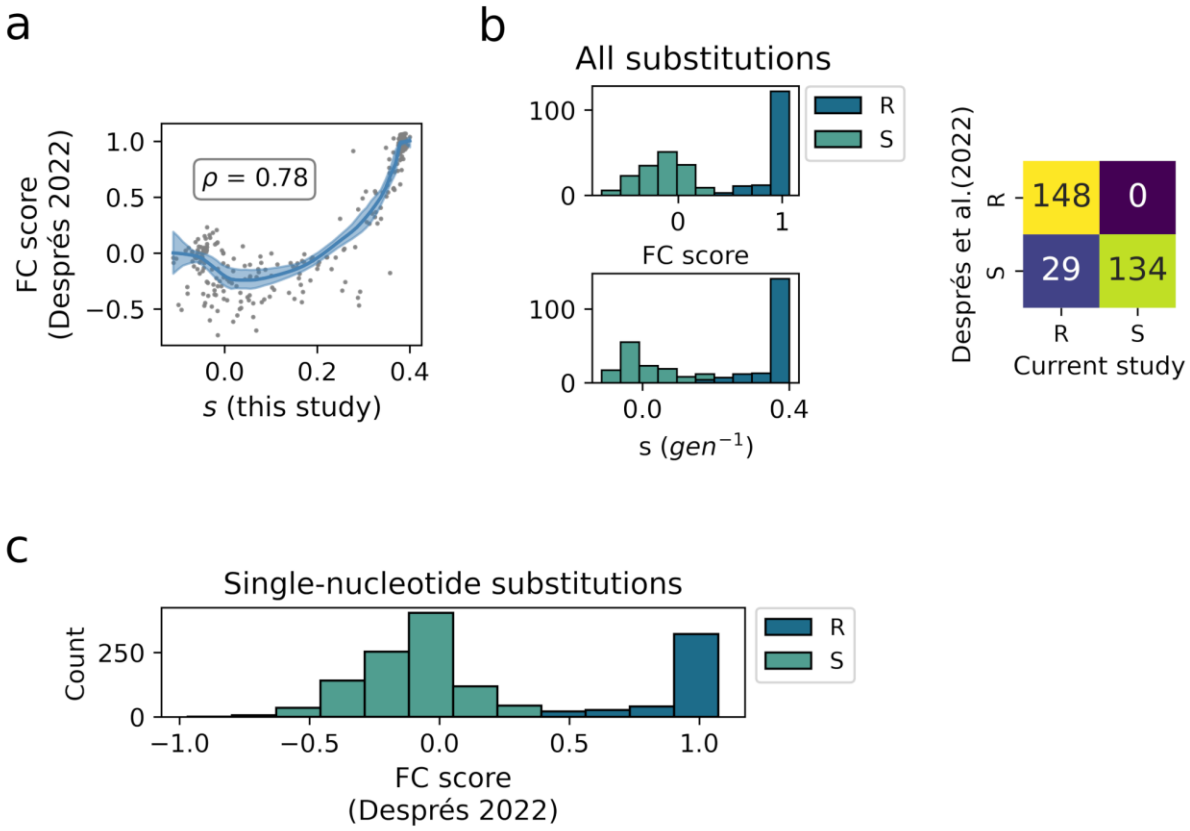
Extended Data Fig. 2: The growth rate of wild-type yeast varies sharply across a range of 5-FC and cytosine concentrations. a, Exponential growth rates for the wild-type reference strain AKD1123 across 5-FC concentrations. Two replicate cultures were assayed for each tested concentration (dots). The line displays a fitted five-parameter logistic function (Methods), while the vertical dashed line identifies the 12 μM concentration used in the bulk competition experiment. b, Same as panel a, but for a range of cytosine concentrations in media lacking uracil. c, The dose-response curve obtained for 5-FC is consistent with a previously published relationship. The curve from a is shown after conversion into relative inhibition levels, using growth rates measured in absence of 5-FC within the same experiment as comparison (0 relative inhibition). This curve, labeled "Exponential - This paper", was used in downstream analyses to obtain the relative inhibition level associated with a given 5-FC concentration. When the same growth curve data is used to compute maximum linear growth rates (rather than exponential growth rates) and then fit a dose-response curve as in Després et al.[31], the published relationship is replicated. d, Same as panel c, but using the cytosine growth data shown in panel b. Here, the published relationship cannot be replicated by using the same analysis approach. The curve, labeled "Exponential - This paper", was nonetheless used in downstream analyses to obtain relative growth levels.



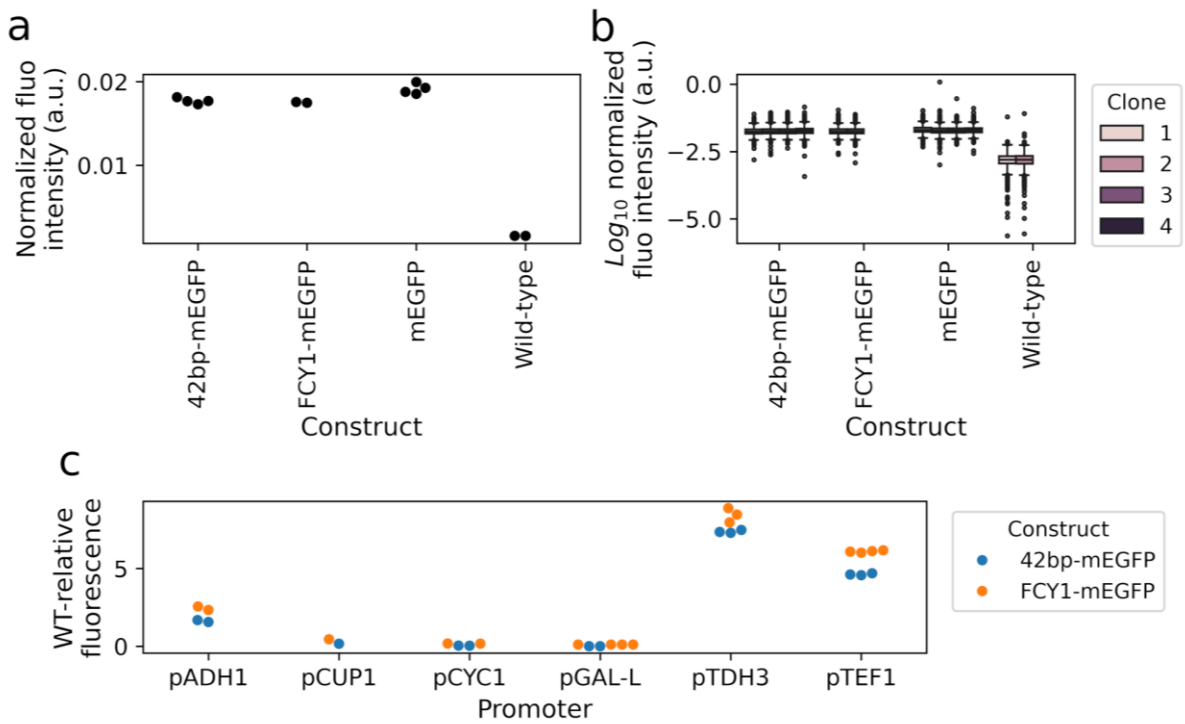
Extended Data Fig. 3: Complete overview of the bulk competition experiment. a, Full experimental design showing all 32 cultures and their relationships as transformation and/or culture replicates. Each pool of substitutions (Sub), insertions (Ins) or deletions (Del) was constructed four times and corresponding replicates were mixed to obtain each of the four *transformation replicates* (grey rectangles). Each transformation replicate was used to inoculate two culture replicates, with or without the addition of a set of previously characterized coding sequence mutations (*FCY1*; n=960)[31]. Created in BioRender. Aubé, S. (2025) https://BioRender.com/acpk3ks. **b,** The selected coding sequence variants cover the full range of variant effects reported in[31], both in 5-FC and cytosine (without uracil) conditions, where Fcy1 activity is respectively toxic and essential.



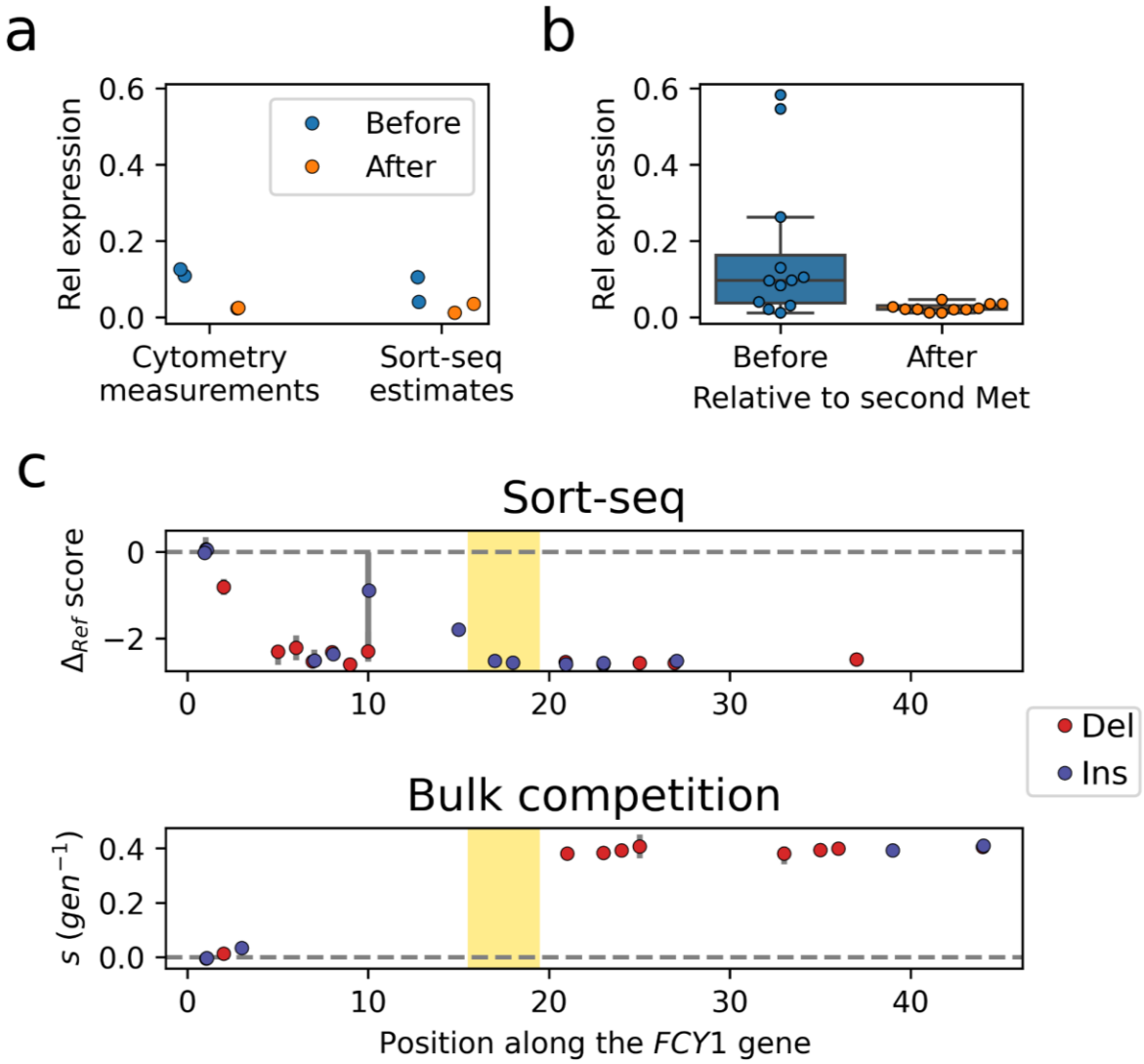
Extended Data Fig. 4: The number and magnitude of small fitness effects observed among promoter single-mutants are consistent with the level of noise in the bulk competition experiment. a, Selection coefficient measurements for the same genotype typically cover an interval of ~0.028 gen^{-1}. The distributions of maximal $|\Delta s|$, the difference between the two most extreme selection coefficient estimates obtained across replicate cultures (n=4), are shown for all assayed single-nucleotide substitutions (Sub; n=462), insertions (Ins; n=462) and deletions (Del; n=96) in the *FCY1* promoter. **b,** The number of mutations exceeding different thresholds of minimal fitness effect is consistent with the expectation derived from random shuffling of the data. Across a range of minimal $|s|$, all genotypes for which the lowest fitness estimate is above $+s$ (significant beneficial effects) or the highest estimate is below $-s$ (significant deleterious effects) are shown (dots). The solid line and shaded area display the median and associated 95% confidence interval for the number of significant effects (using the same definition) obtained from random permutations of the data (n=10 000). Within each iteration, 1020 mock single mutants were assembled by randomly combining selection coefficients from each of the four replicate cultures (Methods). **c,** Twelve promoter single-mutants display $|s| > 0.01$ in the bulk competition experiment. For each genotype, all replicate s estimates are shown (dots). The identifiers of the mutations (x axis) inform on their type (deletion, insertion or substitution), position relative to the *FCY1* coding sequence (where 0 is the base immediately upstream of the gene), and nucleotide change. **d,** The small fitness effects measured are not clearly related to corresponding changes in *FCY1* expression level. For mutants included in panel **c**, the median selection coefficient estimate is compared with WT-relative expression level estimated from sort-seq (see **Fig. 4e**). Whereas the correlation is negative as expected, the second highest fitness effect is observed for a mutation increasing expression (>1) and the most negative selection coefficient is associated with a moderate decrease of *FCY1* expression.



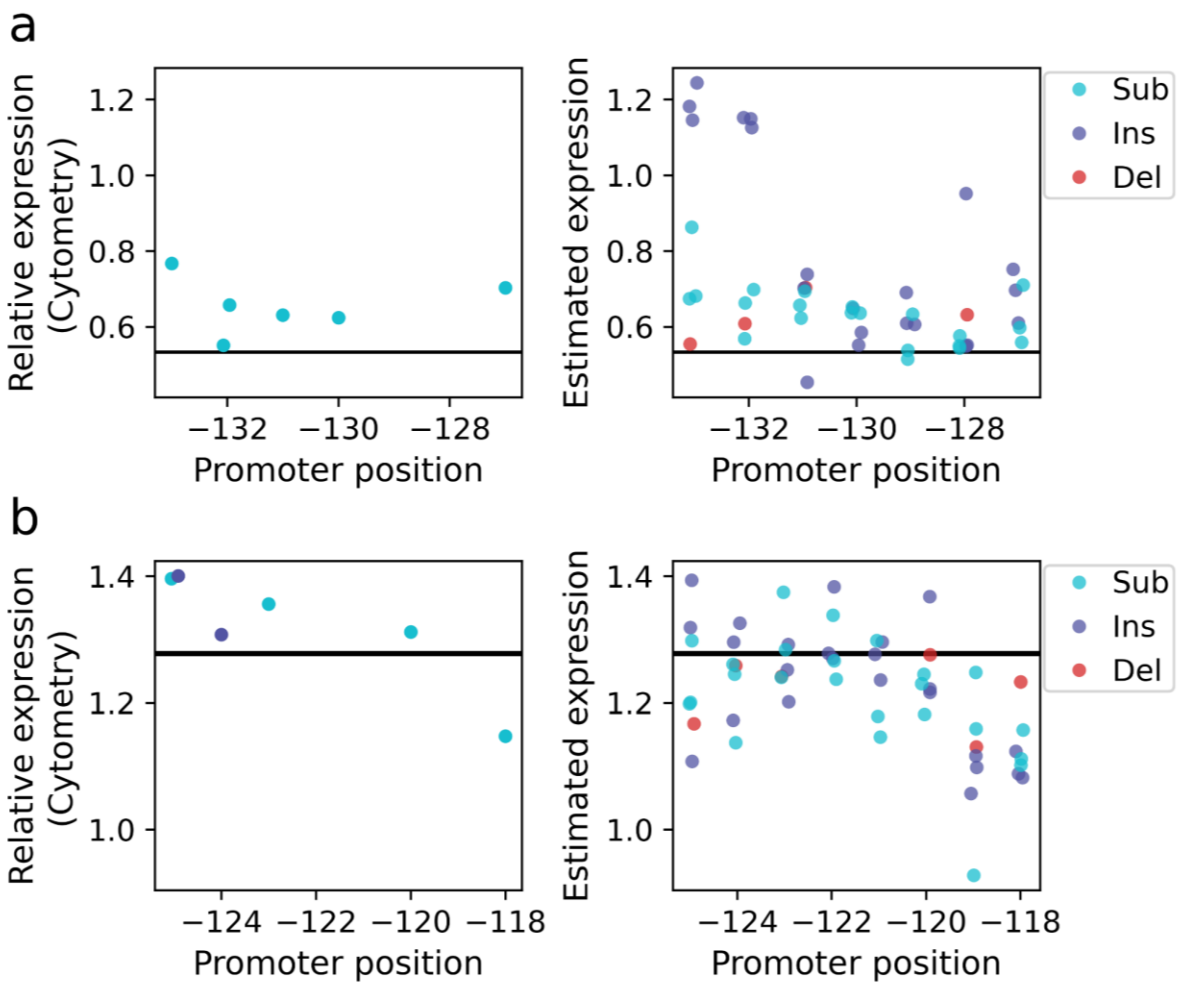
Extended Data Fig. 5: Single-nucleotide substitutions in the *FCY1* coding sequence readily confer resistance to 5-FC, across experiments. **a,** Comparison of median selection coefficients for coding sequence mutants (n=960) added to the bulk competition experiment with previously published variant effect scores measured in the same conditions[31]. Values have been aggregated into medians for all genotypes which result in the same amino acid sequence, such that n=311 points are shown. A LOWESS fit highlights a mostly monotonically increasing trend, as shown by a high Spearman correlation ($p < 10^{-64}$). **b,** The set of coding variants which confer resistance to 5-FC is consistent across the two experiments. All mutants which were added to the "+*FCY1*" samples (**Extended Data Fig. 3a**) were classified into resistant (R), and sensitive (S) variants according to their effect in 5-FC conditions, using Gaussian mixture models (*Methods*). This was done using either the FC score reported by[31] (top) or the selection coefficients inferred in the current study (left). The confusion matrix (right) shows that, among all variants which were classified as resistant using data from[31], none appear sensitive to 5-FC in the current study. As in the previous panel, mutated sequences were collapsed into amino acid substitutions resulting in 311 variants. **c,** Almost a third of single-nucleotide substitutions in *FCY1* confer resistance to 5-FC. The distribution of variant effect scores from[31] for all single-nucleotide substitutions in *the coding sequence*. For each genotype (n=1411), the score reported for the corresponding amino acid variant was used. Only 20 of the possible nucleotide substitutions are missing, mostly in the first and last codons. Variants are identified as resistant (R) or susceptible (S) according to the classification shown in panel **b**. The "R" class includes ~29% (408/1411) of the single substitutions shown.



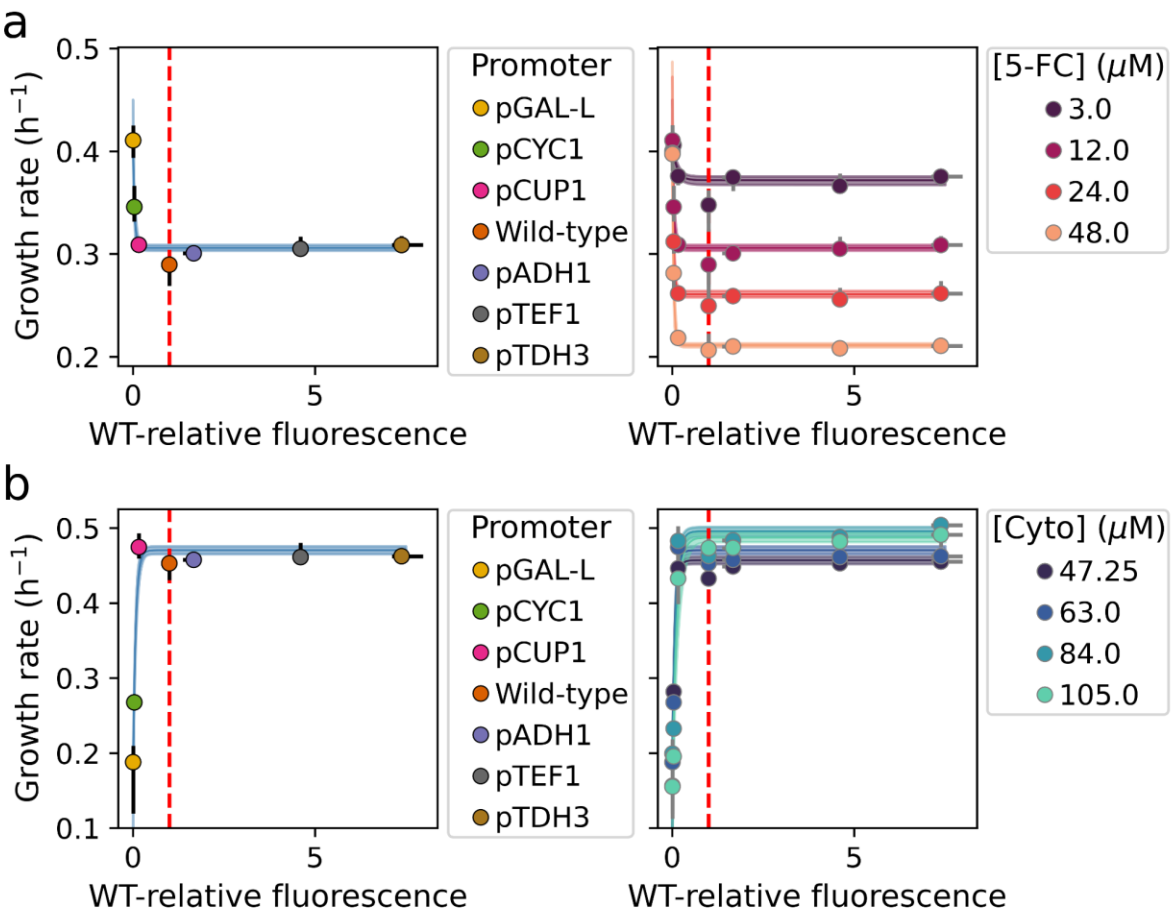
Extended Data Fig. 6: **Validation of the 42bp-mEGFP construction. a,** Medians of normalized fluorescence intensity, as measured by flow cytometry (*Methods*), for three mEGFP insertions at the genomic *FCY1* locus. Each point is a distinct clone of the corresponding strain, except for the WT. In the 42bp-mEGFP construct, which was used in the current study to measure promoter activity, the mEGFP green fluorescent protein is fused in C-terminal of the first 42 bp of the *FCY1* coding sequence. In contrast, FCY1-mEGFP designates a C-terminal fusion with full-length *FCY1,* while mEGFP stands for the insertion of the fluorescent protein in place of the *FCY1* coding sequence. **b,** As on the left, but showing the full distribution of log-scaled normalized fluorescence intensities for each clone (5000 events). **c,** Comparisons of median WT-relative fluorescence intensities for the different promoter constructs, in the 42bp-mEGFP or FCY1-mEGFP backgrounds. In each case, the WT-relative value has been computed relative to the corresponding reference (WT *FCY1* promoter sequence, upstream of either the 42bp-mEGFP or FCY1-mEGFP construct). The ranks of the promoters are unchanged between the two backgrounds, although WT-relative fluorescence intensities are consistently higher when the fluorescent protein is fused to full-length *FCY1*. Values shown for the 42bp-mEGFP are the same ones which were used throughout the paper to draw expression-fitness curves, although replicate measurements obtained from the same clone were aggregated into grand medians. The FCY1-mEGFP strains had not been validated prior to cytometry. All clones which displayed a fluorescence intensity clearly different from the corresponding WT are shown. Because the promoter cassettes were directly inserted in place of the endogenous promoter in the WT background, the WT expression level (and thus fluorescence intensity) is expected in cases where integration failed (*Methods*).



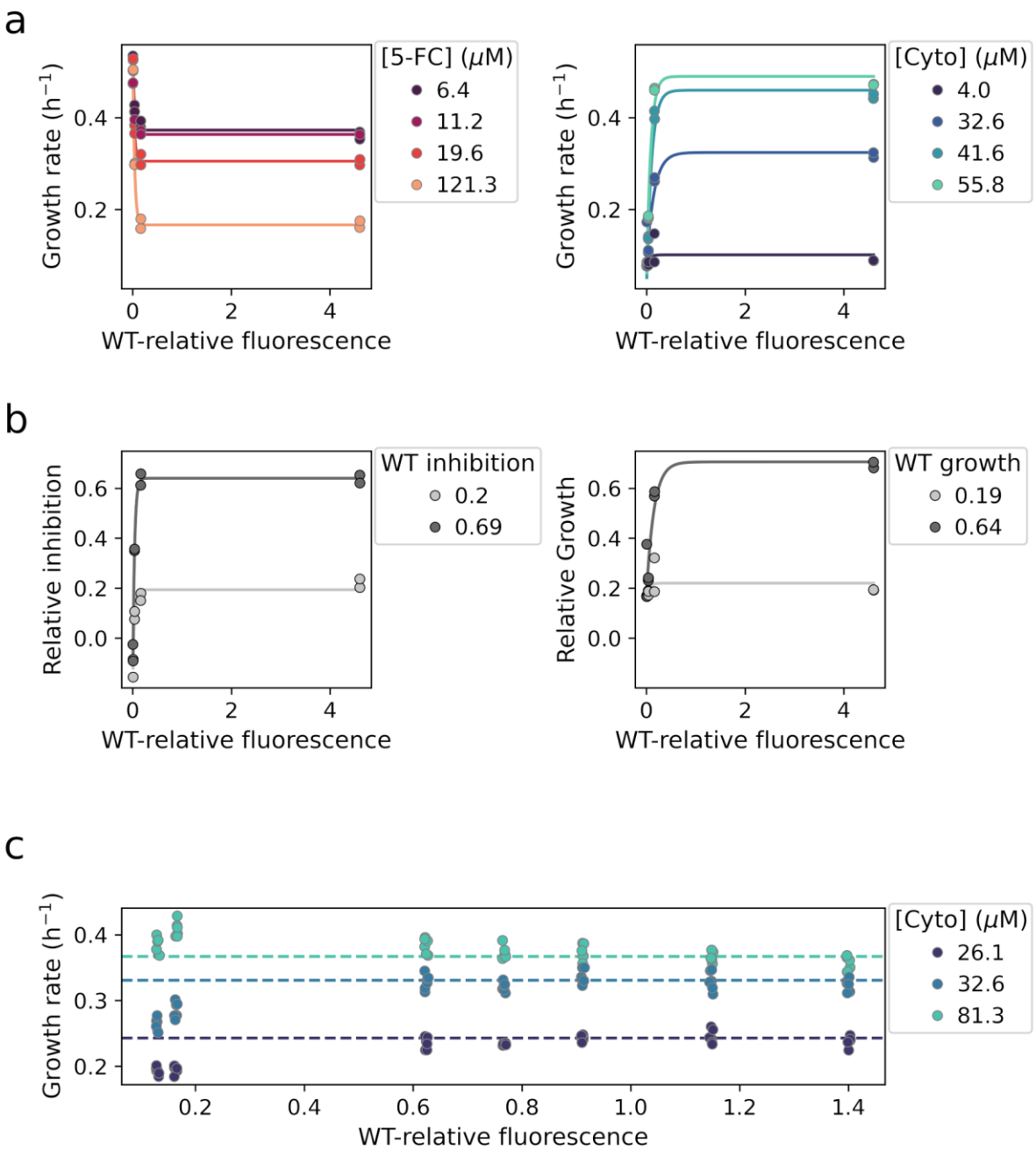
Extended Data Fig. 7: **Indels before and after the second methionine codon of *FCY1* have contrasting effects on expression level. a,** Comparison of relative expression levels for the four indel mutations which were included among the set of validation mutations, classified according to whether they occur before or after the second methionine codon of *FCY1*. Cytometry measurements from reconstructed mutants as well as estimates obtained from the large-scale sort-seq experiment are shown. **b,** Relative expression levels, as estimated from the sort-seq experiment, for all detected indel mutations in the *FCY1* coding sequence. Those occurring after the second methionine are clearly more impactful. **c,** Median effect of insertion and deletion variants detected in the sort-seq (top) and bulk competition (bottom) experiments. The shaded area encompasses the three bases of the second methionine codon. Mutations occurring before this codon were recovered at a much lower frequency from the bulk competition, which is consistent with the observation that they do not completely abolish *FCY1* expression and thus do not confer resistance.



Extended Data Fig. 8: Deletion of putative TF binding sites is not more impactful than single-nucleotide mutations in the same regions. a, Deletion of positions -133 to -127 of the *FCY1* promoter, corresponding to site 1 (positions -132 to -127) of **Fig. 4e**. The WT-relative fluorescence of the deletion mutant, from three replicate measurements in flow cytometry, is shown as an horizontal line whose width displays the minimum and maximum values among the three corresponding medians. Among reconstructed validation mutants (left), one single-nucleotide substitution decreases *FCY1* expression as much as the deletion, as shown by its median WT-relative fluorescence intensities (dots; grand median of three replicate cytometry measurements). Among all mutations assayed by sort-seq (right), the most impactful single-nucleotide changes decrease expression as much as the deletion, as shown by the corresponding relative expression level estimates (dots; median of the 95% prediction interval obtained from sort-seq calibration). **b,** Deletion of positions -125 to -118, corresponding to site 2 (-125 to -120) of **Fig. 4e**. The plots are the same as in **a**, but this time for a deletion which increases *FCY1* expression. Cytometry measurements of validation mutants (left) and sort-seq estimates (right) both show that many single-nucleotide changes have an effect that is comparable to that of the deletion – or even larger.



Extended Data Fig. 9: The insensitivity of fitness to changes in *FCY1* expression level is constant across many concentrations of 5-FC and cytosine. a, Fitness functions in 5-FC. In [12 μM] 5-FC (left), a constant plateau of low fitness is reached starting from ~0.18 WT-relative expression (same as **Fig. 5b**). Across three additional concentrations of 5-FC (right), the boundary of the plateau does not change. Its height however varies, such that lower growth rates are observed for increasing 5-FC concentrations. **b,** Fitness functions in media lacking uracil, where deamination of cytosine by Fcy1p is essential. When cytosine is added at [63 μM] (left), a similar plateau is observed from ~0.18 WT-relative expression, such that any further increase in *FCY1* expression does not increase growth rate. When three additional concentrations of cytosine are tested (right), the boundary of the fitness plateau also remains constant. Error bars around each dot display minima and maxima of three replicate measurements for growth rate and relative fluorescence intensity for each combination of promoter and 5-FC/cytosine concentration. In addition, dashed vertical lines identify the WT expression level (relative fluorescence intensity of 1.0).



Extended Data Fig. 10: The expression-fitness function of *FCY1* displays a sign inversion between 5-FC and cytosine conditions. a, Fitness functions across 5-FC and cytosine concentrations covering the full ranges of the corresponding dose-response curves (**Extended Data Fig. 2**). Exponential growth rate measurements were performed in duplicates for each of the promoter constructs which define the elbow of the fitness function: pGAL-L, pCYC1, pCUP1 and pTEF1. In 5-FC (left), the conditions shown, from the lowest to the highest concentration, respectively result in relative inhibition levels of 0.20, 0.29, 0.40 and 0.69. In cytosine (right), the concentrations shown are respectively associated with the following relative growth levels: 0.19, 0.64, 0.93 and 0.97. **b,** Comparison of fitness functions obtained in 5-FC (left) and cytosine (right**)** concentrations resulting in matching levels of relative inhibition or growth for the WT reference strain (**Extended Data Fig. 2**). Although growth does not reach 0 in cytosine conditions when *FCY1* expression is decreased to very low levels, the two pairs of curves display very similar shapes. Same data as in panel **a**, but converted into relative inhibition or growth values using the relevant dose-response curve. **c,** Increasing *FCY1* expression through single mutations in the promoter does not provide a benefit in cytosine conditions,

where Fcy1 activity is essential for growth. Exponential growth rate measurements were performed in triplicates for eight reconstructed promoter mutants covering the full range of expression changes available from single-nucleotide changes (**Fig. 5d**). From the lowest to the highest, the three cytosine concentrations used respectively result in relative growth levels of 0.42, 0.64 and 0.97. The dashed horizontal lines display the medians of growth rates obtained for the two mutants closest to the WT expression level. Across all three concentrations, the mutant associated with the highest expression level is not clearly fitter than this baseline. This confirms that, just like they are not deleterious in 5-FC conditions, promoter mutations increasing *FCY1* expression are not beneficial in cytosine conditions.