Adaptation is less accessible through mutations in promoters than in coding sequences when large effect sizes are needed

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

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

Abstract

Mutations within promoters and coding sequences are both involved in adaptation, but their relative contributions remain to be compared directly. Using the fungal enzyme cytosine deaminase, we examine if adaptation to 5-fluorocytosine (5-FC), which requires reduced cytosine deamination, can be reached by single promoter mutations. We generated all single-nucleotide substitutions and indels in the yeast *FCY1* promoter and assayed the resulting mutants in presence of 5-FC. This revealed that no promoter mutation is sufficient for adaptation to occur. In contrast, 29% of single-nucleotide substitutions in the *FCY1* gene have previously been reported as conferring adaptation to 5-FC. We hypothesized that this discrepancy could arise if *FCY1* expression was robust to promoter mutations or if fitness was insensitive to expression level changes. Combining large-scale expression measurements with the experimental characterization of the corresponding expression-fitness function, we found that, while 24% of promoter mutations significantly affect expression, the shape of this function prevents them from being adaptive. Because this relationship is step-like, the curve is flat around wild-type level and adaptation can only emerge from a severe reduction of expression. These results highlight how adaptive phenotypes may be more mutationally accessible in coding sequences than promoters, especially when large effects are beneficial.

Introduction

For adaptation to occur, advantageous phenotypes must be accessible through mutations. This largely depends on the underlying fitness landscape or genotype-phenotype map [1,2], defined by the respective fitness of all possible genotypes. Although evolution could occur through any path created by single mutations between neighboring sequences, only immediate fitness increases are favored by natural selection. Adaptation is therefore mostly accessible through trajectories where each successive mutation is beneficial [3,4]. For this reason, any state of higher fitness will be difficult to reach if no single mutant from a starting genotype is adaptive (**Fig. 1a**). Whether beneficial mutations are immediately available thus largely determines the adaptive potential of a population of organisms.

A powerful approach to assess the accessibility of adaptation is hence to perform saturation mutagenesis, through which the effect of all single mutations in a sequence can be mapped. When focused on protein-coding sequences, such work is known as Deep Mutational Scanning or DMS [5]. This type of experiment is especially effective when an adaptive phenotype can readily be measured in the laboratory, as is the case for resistance to drugs such as antimicrobials. Multiple DMS studies have accordingly shown that resistance is accessible through single substitutions in the coding sequences of drug targets [6–11]. This work may however represent an incomplete sampling of the mutational paths that lead to adaptation, since mutations in the regulatory elements of genes can also be impactful (**Fig. 1b**). Multiple reports of adaptive mutations linked to changes in gene expression or gene dosage [12–18], as well as the identification of clear relationships between fitness and the expression level of many genes [16,19–24], indeed demonstrate that regulatory mutations can contribute to adaptation.

Numerous studies have investigated the effect of mutations within promoters and cis-regulatory sequences on expression levels using random or saturation mutagenesis, for instance in bacteria[25], yeast [26–28], *Drosophila* [29] and mammals [30–33]. Yet, much less is known about the fitness impact of mutations within promoters, especially in comparison with the effects of coding mutations in the corresponding genes. The relative accessibility of adaptive phenotypes through such changes thus remains to be assessed thoroughly (**Fig. 1c**). Even in few cases where the fitness consequences of promoter mutations are well known, for instance in *TDH3* [16,34] and *SUL1* [14] in yeast, only a handful of coding sequence mutations have comparatively been assayed [35,36].

To compare the relative contributions of promoter and coding sequence changes to adaptation, we focused on the adaptive loss of Fcy1 activity in yeast exposed to prodrug 5-fluorocytosine (5-FC). While this enzyme's canonical activity is cytosine deamination, it also converts 5-FC into 5-fluorouracil, which leads to growth arrest and cell death [37]. Accordingly, complete or partial *FCY1* loss-of-function (LOF) – which is readily available through single-nucleotide mutations within the coding sequence [6] – confers resistance and is adaptive. To estimate whether promoter mutations are similarly impactful, we performed saturation mutagenesis of the *FCY1* promoter. This experiment revealed no measurable fitness gain in medium with 5-FC, although a quarter of single-nucleotide changes significantly affect gene expression level.

Through the characterization of the underlying expression-fitness function, we show that this discrepancy emerges from the shape of this relationship. Because this fitness function is largely flat around the endogenous expression level, only coding sequence changes can sufficiently decrease the amount of functional *Fcy1* to confer resistance. These findings highlight how adaptive phenotypes may be less accessible through mutations within promoters, at least when large changes to the underlying molecular traits are beneficial.

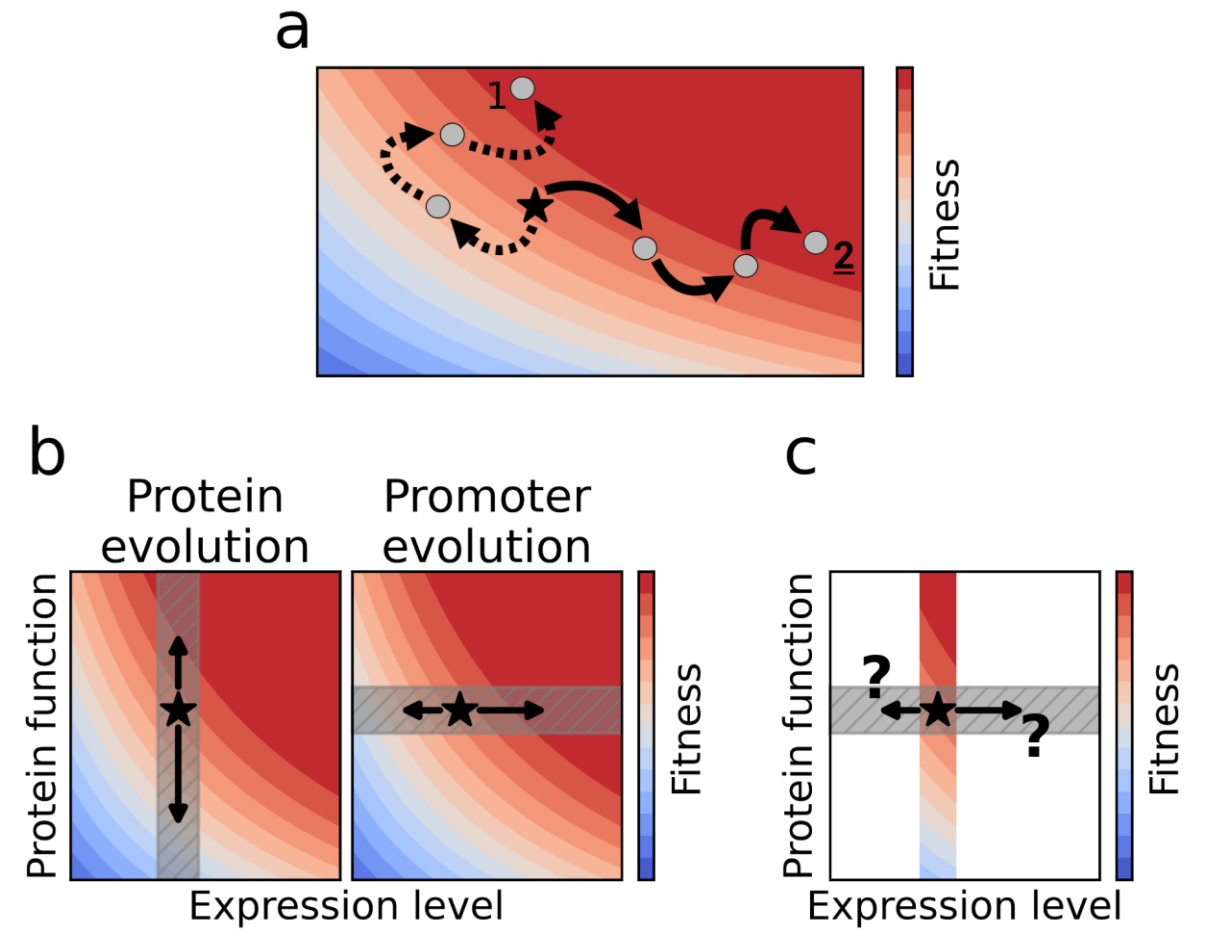


Fig. 1: Adaptation can occur through mutations in coding and regulatory sequences. a, Not all adaptative routes are equally accessible. Two equally advantageous genotypes (1 and 2) are shown on a hypothetical fitness landscape. Both require the acquisition of three mutations (dots) from the wild-type (WT) genotype (star), but only one of the mutational trajectories is accessible. Because it first involves the fixation of a mildly deleterious mutation, mutational path 1 is not favored by natural selection and hence difficult to reach from the chosen starting point, as illustrated by dashed lines. This illustrates the importance of characterizing the effects of single mutations from a WT genotype to study its adaptive potential. **b,** Comparison of protein and promoter evolution, over a landscape defined by expression level and protein function. Single mutations in the coding sequence (left) affect protein function (stability, activity, etc), while single mutations in the promoter modify gene expression level and thus protein abundance. As such, they explore two complementary slices of the landscape. Saturation mutagenesis experiments on a gene and its promoter respectively study the same two slices, at WT expression level or WT protein function. Such studies can thus assess whether high fitness regions can be reached through each of the two types of mutations. **c,** The current knowledge of fitness landscapes is incomplete. Since systematic studies of the fitness effects of mutations have mostly explored protein function, only a small region corresponding to WT

expression level is known in most cases. The orthogonal region is unknown, such that the accessibility of adaptive (high-fitness) mutations in promoters remains to be assessed thoroughly. Here, we examine the fitness effects of regulatory mutations for a gene whose fitness landscape has previously been studied in the protein dimension.

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 first 150 bp [38], 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 performed 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* [39,40]). 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 of the 154 nucleotide positions, the wild-type (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 and without 5-FC, resulting in 16 cultures. The prodrug was added at a final concentration of 12 μM, which has been shown to inhibit the growth of *Saccharomyces cerevisiae* by ~50% and was used in the previous DMS of the *FCY1* coding sequence [6] (**Supplementary Fig. 3**). Additional replicates of all 16 cultures were also done by combining the promoter mutants with 960 previously characterized coding sequence variants[6], to ensure that the fitness effects of the two types of mutations could be directly compared (**Extended Data Fig. 2**).

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 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 [41] (**Fig. 2c**).

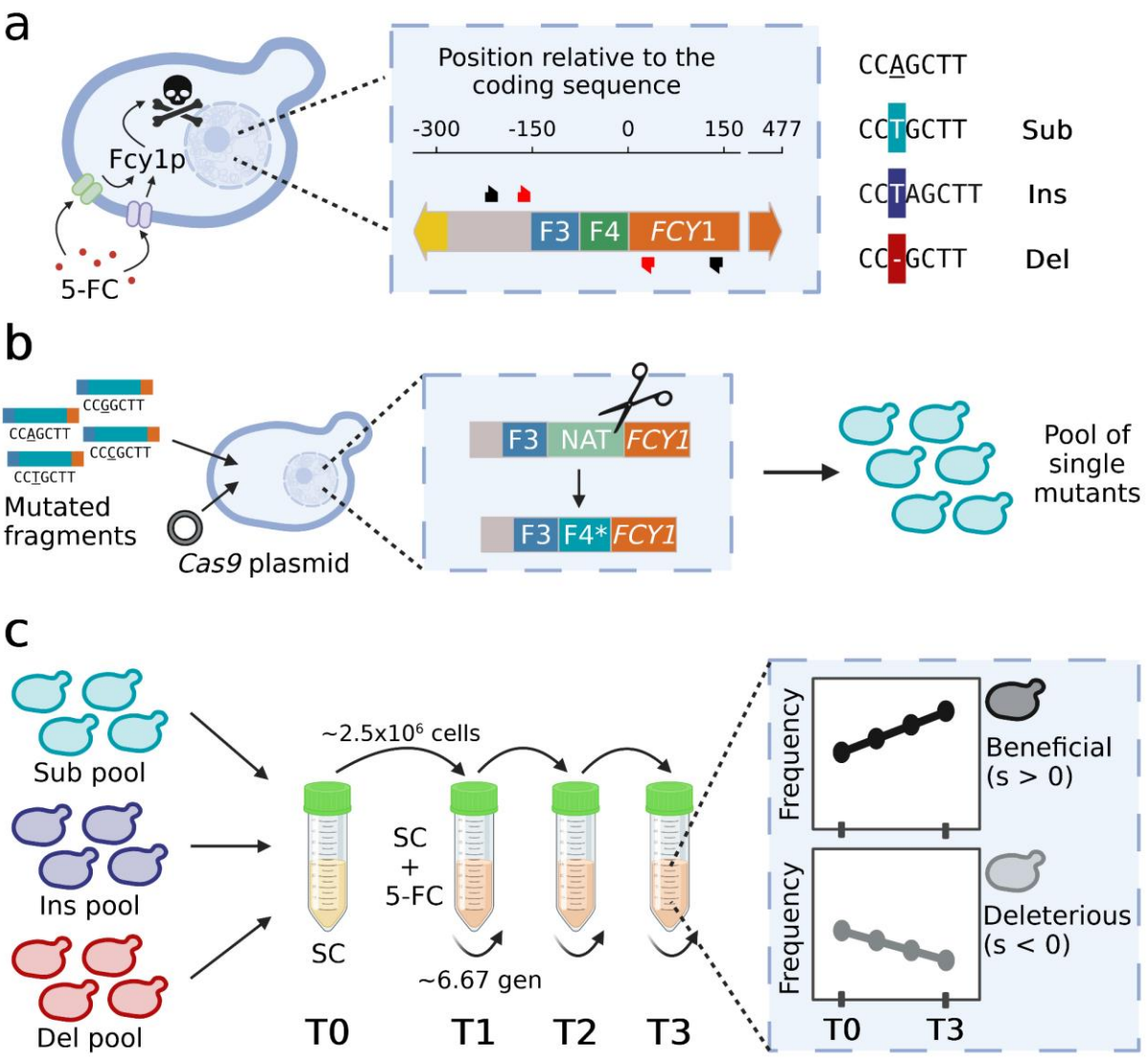


Fig. 2: Experimental design of the *FCY1* promoter deep mutational scanning. a, 5-FC is converted into 5-FU, a toxic compound, by the cytosine deaminase (Fcy1p), such that *FCY1* expression can have a strong effect on fitness. Fragments 3 (F3) and 4 (F4) of the *FCY1* promoter were replaced with a library 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.

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, but sensitive to coding mutations

After correcting for the impact of secondary mutations, the inferred selection coefficients for the F3-F4 mutants revealed only small fitness effects (**Fig. 3a**). Across all assayed promoter mutations, the median *s* estimates obtained from replicate experiments ranged from ~-0.06 to ~0.06 gen^{-1}. Experimental noise was moreover of a similar magnitude as these deviations from 0, as shown by an average difference of ~0.04 gen^{-1} between the highest and lowest *s* among replicate estimates for the same genotype, up to ~0.18 gen^{-1} (**Supplementary Table 8**). A small number of secondary mutations remained undetected (not found within the longer sequencing fragment), resulting in outlier selection coefficients for one pair of culture replicates within some F3-F4 genotypes. Overall, following a stringent removal of these rare outliers (**Supplementary Fig. 9**), none of the 1020 *s* estimates for *FCY1* promoter mutations appeared significantly different from 0 ($p > 0.05$; one-sample t-test with Benjamini-Hochberg correction at FDR=5%).

This lack of fitness effects is not due to a failure of the competition experiment, as confirmed by two layers of controls. 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. 10**). More importantly, the fitness inferred for the added coding sequence variants replicate previous findings on the effect of amino acid substitutions in Fcy1p [6] (**Fig. 3b**). There is in addition almost perfect agreement between the two experiments regarding the number and identity of resistance-conferring mutations (**Extended Data Fig. 3**). Although samples which contained coding sequence variants also included the complete pool of promoter mutations (**Extended Data Fig. 2**), the resulting cultures were not used to precisely estimate F3-F4 mutational effects. This is because the numerous resistant genotypes within the coding sequence pool outcompeted promoter mutants and rapidly took over the population (**Extended Data Fig. 4a**), thus decreasing the resolution of F3-F4 fitness estimates. Although much

noisier, the corresponding selection coefficients were qualitatively similar to those reported in **Fig. 3A** (**Extended Data Fig. 4b**).

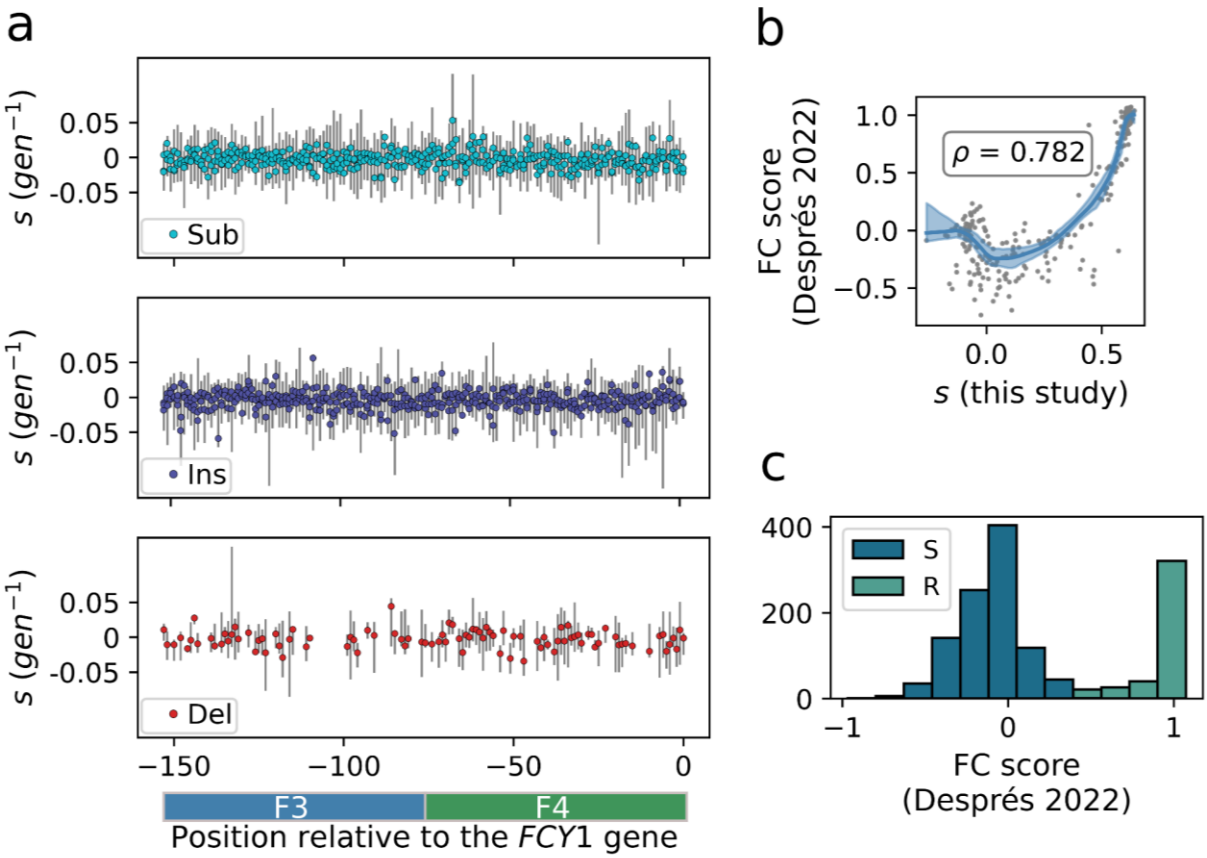


Fig. 3: Mutations in the *FCY1* promoter do not have a measurable impact on 5-FC resistance. a, Fitness effects of single-nucleotide mutations in the *FCY1* promoter in 5-FC conditions. All substitutions (Sub; n=462), insertions (Ins; n=462) and deletions (Del; n=96) are shown. Points display the median 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. 7**). **b,** Comparison of median selection coefficients for *FCY1* coding sequence mutants added to our experiment with previously published variant effect scores measured in 5-FC [6]. 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 ($p < 10^{-64}$). **c,** Distribution of variant effect scores obtained in 5-FC [6] for all single-nucleotide substitutions in *FCY1*. For each genotype (n=1411), the score reported for the corresponding amino acid sequence was used [6]. Only 20 of the possible nucleotide substitutions are missing, mostly in the first and last codons. Variants which are susceptible and resistant to 5-FC are respectively identified with "S" and "R". About 29% are classified as resistant (*Methods*).

Negative and positive controls thus confirm that 5-FC selected against cytosine deaminase activity, as expected, and that the absence of measurable fitness effects is not artefactual. This

weak impact of promoter mutations contrasts markedly with the strong effects of mutations in the *FCY1* coding sequence. Within this gene, ~29% (408/1411) of single-nucleotide substitutions indeed result in codon changes which confer resistance to 5-FC in the same conditions (**Fig. 3c**)[6]. Moreover only 67 of these 408 mutations introduce a premature stop codon. Notably, this comparison underestimates the accessibility of adaptation through mutations in the *FCY1* gene. Most indels in this coding sequence, which were not assayed, are indeed likely to grant resistance by introducing frameshifts.

Within our *FCY1* model system, single-nucleotide mutations in the promoter thus appear to provide little to no paths to adaptation, contrary to changes in the corresponding coding sequence. Such weak fitness effects of promoter mutations could emerge through two mechanisms. First, the abundance of Fcy1p could be robust to changes in its promoter. Second, fitness in 5-FC conditions could be insensitive to variations in *FCY1* expression level, such that no mutational effects would be measurable through bulk competition assays. Discriminating between these two competing hypotheses first requires measuring the impact of *FCY1* promoter mutations on expression level.

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[42,43]. 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. 5**). 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**). 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 in two culture replicates).

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}, not all completely abolished 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. 6 a,b**). This is consistent with a previous report of a minor translation initiation site at this codon [44].

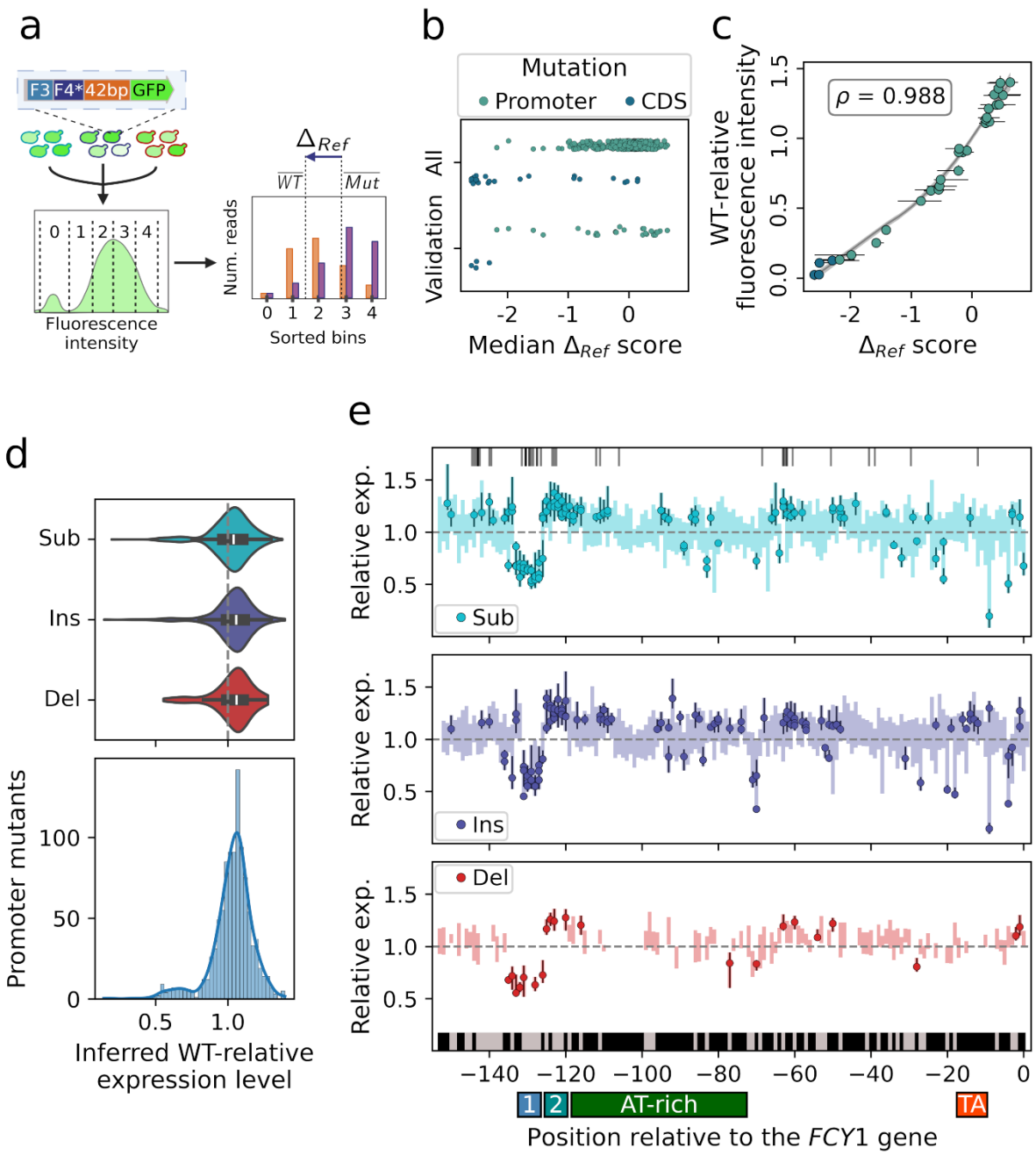


Fig. 4: Mutations in the *FYC1* promoter impact expression level. a, Schematics of the sort-seq experiment. Yeast cells from pools of promoter mutants were sorted into five bins according to their fluorescence intensity (representative distribution shown). From the sequencing of each bin, WT-relative fluorescence intensity could be estimated. 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

292 and the corresponding shaded area display a LOWESS fit with its 95% confidence interval
293 (*Methods*). **d,** Distributions of WT-relative expression levels for all single-nucleotide promoter
294 mutations, as inferred from the calibration of sort-seq scores. At the top, distributions are shown
295 separately for substitutions (Sub; n=461), insertions (Ins; n=460) and deletions (Del; n=96),
296 with a dashed line indicating WT expression level (1.0). At the bottom, the full distribution,
297 combining all three types of single mutants, is shown (n=1019). **e,** Distribution of promoter
298 mutations with a significant effect on expression (n=245), along the mutagenized region.
299 Separately for Sub (n=113), Ins (n=111) and Del (n=21) variants, the estimated WT-relative
300 expression (point) and its corresponding 95% prediction interval (error bar) are shown for each
301 mutant affecting FCY1 expression by 5% or more (minimum or maximum of the corresponding
302 95% prediction interval; *Methods*). Shaded rectangles display the minima and maxima across
303 all prediction intervals for each position along the sequence, including variants which are not
304 shown. In the upper plot, vertical lines show the midpoint of hits (n=41) to known TF
305 recognition motifs, identified using YeTFaSCo [45]. In the lower plot, the black and white
306 rectangles respectively display A/T and G/C positions. Promoter positions corresponding with
307 the two main clusters (sites "1" and "2"), the approximate AT-rich region and the TATA-like
308 element ("TA") [46,47] are also identified.
309
310 To validate the sort-seq assays, we selected 30 mutants for independent reconstruction (**Fig.**
311 **4b**). For each newly constructed strain as well as for the WT reference, fluorescence intensity
312 was measured by flow cytometry (*Methods*). The resulting WT-relative intensities are in strong
313 agreement with the corresponding median Δ_{Ref} scores (**Fig. 4c, Extended Data Fig. 7**). We
314 additionally used this validation to draw up a calibration curve, enabling the conversion of sort-
315 seq scores into more easily interpretable relative fluorescence intensities. To this end, we fit a
316 LOWESS regression (**Fig. 4c**), to avoid making specific assumptions about the relationship
317 between the two variables. This was performed within a bootstrapping routine which involved
318 repeated resampling of Δ_{Ref} values as well as validation fluorescence measurements, thus
319 generating 95% prediction intervals for the curve itself and for the relative fluorescence
320 intensities of all genotypes (*Methods*, **Supplementary Table 9**). We hereafter refer to the
321 medians of the latter intervals as WT-relative expression levels for the corresponding
322 mutations, where a value of 1 denotes the WT level.
323
324 Examining the inferred relative expression levels for all F3-F4 mutants reveals a slight bias
325 towards expression increases, with medians lying between 1.04 and 1.06 (**Fig. 4d**). That
326 mutations are more likely to increase expression than to decrease it may appear surprising,
327 since Fcy1p is already highly expressed and is among the 10% most abundant proteins in *S.*
328 *cerevisiae* [48]. While this bias could be artefactual, the y-intercept of the calibration curve
329 suggests otherwise (**Fig. 4c**). It is very close to 1 (~1.00, with 95% prediction interval from
330 0.98 to 1.02) – as expected for a Δ_{Ref} of 0 –, although no WT measurements were used in the
331 LOWESS fitting.
332
333 The overall distribution of effects highlights that most single mutations in the *FCY1* promoter
334 weakly affect expression levels, with 75% of variants lying between values of 0.98 and 1.11
335 (**Fig. 4d**). Yet, we also observe a small number of mutations with larger effects, decreasing
336 expression to ~50% of its WT level, or even lower. This tail appears more pronounced than

337 previously reported in yeast [26,27], potentially because we specifically mutated the most proximal
338 part of the *FCY1* promoter, where mutations are more likely to disrupt processes such as
339 transcription initiation. Among the 1019 F3-F4 mutations, 245 significantly affect *FCY1*
340 expression, by 5% or more (95% prediction intervals). In agreement with a previous study of a
341 yeast promoter [14], these significant mutations are not distributed uniformly along the
342 mutagenized promoter region (**Fig. 4e**). This distribution reveals important determinants of
343 *FCY1* expression level.
344
345 The fact that significant mutations form two main clusters, from positions -132 to -127 (site 1)
346 and from -125 to -120 (site 2), first highlights the importance of transcription factor (TF)
347 binding. Within these two sites, 92.4 % (73/79) of all variants change expression by 5% or
348 more, in a consistent direction within each cluster. Both sites are notably within a 80 bp
349 segment known to be critical for TF binding in yeast promoters [28,42,49], while the whole -132 to
350 -120 region is additionally a hotspot of predicted TF binding sites [45] (**Fig. 4e top;**
351 **Supplementary Table 10**). As such, the clusters likely reveal TF binding sites. Interestingly,
352 the deletion of either of the two sites is not more impactful than the strongest mutations within
353 the corresponding cluster (**Extended Data Fig. 8**), suggesting that single mutations are
354 sufficient to disrupt them.
355
356 The position of the two most impactful mutations additionally highlights how alterations of
357 transcription initiation may severely reduce *FCY1* expression. These variants, a substitution
358 and an insertion which respectively decrease expression to ∼16% and ∼13% of its WT level
359 (cytometry; **Fig. 4c**), both occur at position -9 (**Fig. 4e**). This position happens to be
360 immediately downstream of the previously identified TATA-like element of the *FCY1*
361 promoter [46,47] – from -17 to -10 –, where the transcription pre-initiation complex assembles.
362 The corresponding mutational effects might thus reflect the disruption of transcription
363 initiation. The conflicting report of a most abundant transcript isoform beginning further
364 upstream at -29 [50], however makes the underlying mechanism unclear.
365
366 Finally, the complete set of significant mutations underscores that most single-nucleotide
367 changes within the poly(dA:dT) region of the *FCY1* promoter weakly affect its transcriptional
368 activity. The corresponding AT-rich sequence, approximately between positions -118 and -73
369 (17.39% GC; **Fig. 4e**, bottom), is indeed comparatively depleted in strong mutational effects.
370 Only 17.6 % (52/295) of variants within it are significant, compared to 26.7 % (193/724) in the
371 remainder of the mutagenized region. This may seem contradictory, considering that such
372 regions are important for the establishment of nucleosome-depleted regions within promoters
373 [51,52], which favor transcription and can be sufficient to induce expression from a downstream
374 core promoter [53]. It is however consistent with previous work on the *URA3* promoter, which
375 showed that deletion of more than half of the corresponding 50 bp poly(dA:dT) was necessary
376 to abolish expression [54].
377
378 Overall, we show that 24% (245/1019) of single-nucleotide mutations in the *FCY1* promoter
379 have a significant effect on expression level, many of which appear to disrupt the binding of

380 TFs at two putative sites. As such, *FCY1* expression is sensitive to mutations, meaning that
381 robustness cannot explain the lack of fitness effects in 5-FC conditions.
382

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

384 To investigate the mechanism underlying the absence of measurable fitness effects, we set out
385 to characterize experimentally the expression-fitness relationship for *FCY1* (**Fig. 5a**). Six yeast
386 promoters covering a wide range of expression levels were used to construct two sets of strains,
387 in which each promoter was inserted upstream of either the native *FCY1* sequence or the
388 previously described 42bp-mEGFP construct. The *FCY1* strains were grown in media with 5-
389 FC (12 μM; like in the competition experiment) – to obtain growth rate measurements –, while
390 mEGFP strains were used to measure expression levels (*Methods*). For each promoter, the
391 measurements from the two strains were combined to draw the corresponding expression-
392 fitness curve (**Fig. 5a**).
393
394 This experiment revealed a very steep fitness function, where growth rate decreases sharply
395 until *FCY1* expression reaches ~18% of its WT level and then plateaus (**Fig. 5b**). Consequently,
396 this fitness curve is flat around the WT, such that small expression changes have no fitness
397 effect. The shape of this function is thus consistent with the fact that no single-nucleotide
398 promoter mutation conferred resistance to 5-FC in our bulk competition experiment.
399 Interestingly, varying the concentration of the prodrug does not impact the sensitivity of fitness
400 to changes in *FCY1* expression. Across 5-FC concentrations resulting in ~25% to ~75%
401 inhibition for the WT, growth rate variation is restricted to the same narrow range of relative
402 expression levels and only the height of the corresponding plateau varies (**Extended Data Fig.
403 9a**). The same phenomenon is observed in media lacking uracil, where the deamination of
404 cytosine into uracil by Fcy1p is essential for growth. Across multiple cytosine concentrations,
405 growth rate increases rapidly until ~18% of WT expression is reached and then remains
406 constant as expression is increased further (**Extended Data Fig. 9b**). The expression-fitness
407 function in 5-FC therefore also reflects the canonical deamination function of the Fcy1 enzyme.
408
409 To increase the resolution of the experimental fitness function around WT expression, we
410 leveraged previously characterized validation mutants. The set of alleles used in the calibration
411 of sort-seq measurements, which also includes a small number of indels in the coding sequence,
412 was reconstructed in the native *FCY1* background and the resulting strains were phenotyped in
413 5-FC conditions (*Methods*). In combination with the corresponding relative fluorescence
414 intensities, these growth measurements confirm the flatness of the curve around WT expression
415 level, with all promoter mutants displaying WT-like growth rates (**Fig. 5c**). Among three
416 assayed coding sequence mutations, the only one which reached the high fitness region is
417 associated with a ~98% decrease in fluorescence intensity (**Supplementary Table 11**). This
418 highlights how resistance – and thus adaptation – is only possible through large reductions in

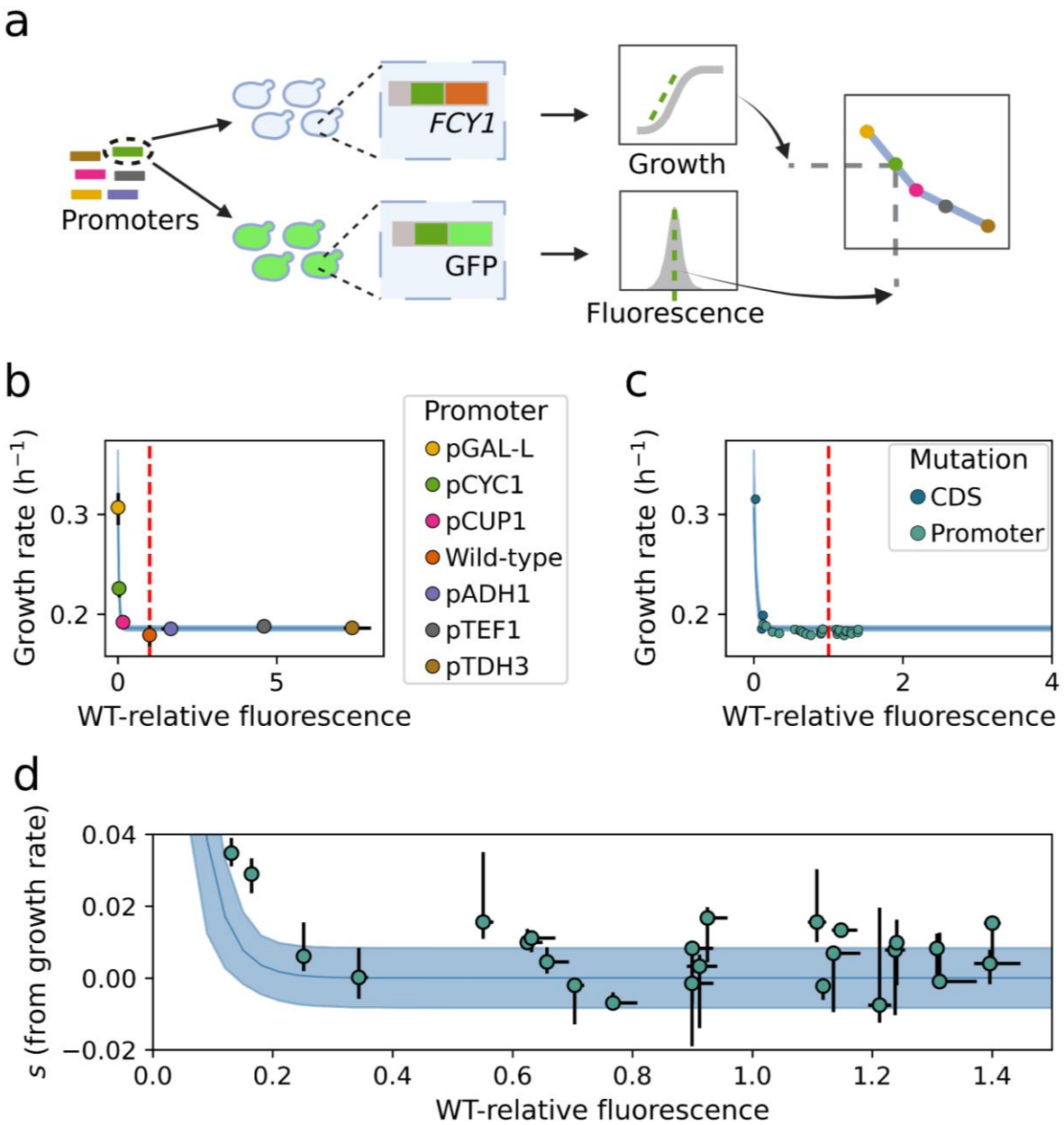


Fig. 5: The *FCY1* expression-fitness function is mostly flat, making fitness robust to promoter mutations. a, Strains with varying expression levels were constructed to measure the fitness function. A set of yeast promoters covering a wide range of activities were introduced in the genome in place of 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*). **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=25) and indels in the first 42 bp of the gene ("CDS"; n=3). The dashed line identifies the WT expression level (1.0). **d,** Reconstructed promoter mutations are mostly neutral. For the same set of single mutations in the promoter (n=25), 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.

Fcy1 activity, in this case due to an almost complete loss of protein expression. Zooming in on the promoter mutants further illustrates this point. Among them, the two single mutations with the largest effect barely reach the threshold below which fitness rapidly increases, although they respectively decrease *FCY1* expression to ∼16% and ∼13% of its WT level (**Fig. 5d**). These two mutants nonetheless display a slight growth advantage, corresponding to selection coefficients on the order of ∼0.03 (**Fig. 5d**) after conversion (*Methods* [55]), which is near the detection limit of our bulk competition assay (**Fig. 3** and accompanying text). Even the two most fit promoter variants are thus readily outcompeted by the many resistance-conferring mutations which can emerge in the coding sequence, associated with selection coefficients over 0.5 (**Extended Data Fig. 6c**). As such, the shape of the fitness function, combined with the fact that no single-nucleotide change in the promoter completely abolishes expression, severely restricts the accessibility of adaptation through cis-regulatory mutations in this *FCY1* system.

Discussion

Adaptive mutations can in principle occur in coding and regulatory regions but the contribution of each remains difficult to estimate [56]. One powerful approach to do so is to focus on a trait whose relationship with fitness is known and systematically assay the adaptiveness of all mutations in both the coding and regulatory regions. To this end, we focused on the promoter of the *FCY1* gene, a protein-coding region within which single-nucleotide changes have been shown to confer resistance to the antifungal 5-FC [6,57]. We surprisingly show that no promoter mutation has a measurable fitness effect in yeast exposed to this molecule, although large changes in expression level (reaching -87%) are observed. In this system, adaptive phenotypes are thus much more readily available through nucleotide variation in the coding sequence than within the corresponding promoter.

In contradiction with this result, our initial data instead suggested that most mutations were highly deleterious. Our investigations revealed that this 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, since this genotype was generated at high frequency within the mutant pool. This comparatively made all true 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 [58]. Such artifacts may be common in large-scale genome editing experiments, due to the increased frequency of single-

478 nucleotide mutations – and especially indels – in synthetic DNA oligonucleotides [59]. This
479 highlights the importance of designing library construction workflows which minimize the use
480 of genic regions as homology arms to guide recombination. Alternatively, these segments
481 should at least be fully included in the genotyping region, as we did to correct our experiment.
482
483 In this study, we observed a striking contrast between promoter and coding sequence mutations,
484 with none of the former conferring resistance, while 30% of the latter do so [6]. This discrepancy
485 can be explained from the shape of the corresponding fitness function, which is mostly flat, but
486 displays steep variation below 18% of the WT amount of functional protein. Due to the step-
487 like nature of this relationship, most promoter mutations have no impact on fitness. While many
488 similar expression-fitness curves have been reported in yeast [20,22], functions where small
489 variations around the WT are immediately impactful are also observed [21,22]. If the abundance-
490 fitness relationship for *FCY1* was instead in the latter category, many promoter mutations
491 would also be able to reach the state of higher fitness. As such, a different shape of fitness
492 function could result in more equal proportions of adaptive mutations in the gene and its
493 promoter. Yet, there is an additional factor underlying the reduced accessibility of adaptation
494 through changes in the *FCY1* promoter: that mutations in the corresponding coding sequence
495 have much larger effects at the molecular level. In accordance with the fitness function we
496 report, the many resistance-conferring mutations in the *FCY1* gene indeed suggest that almost
497 a third of substitutions in this sequence decrease protein abundance and/or activity below 5%
498 of its WT level. In contrast, only two promoter mutations – among more than 1000 – approach
499 the low expression threshold below which fitness begins to increase. Whether coding or
500 regulatory mutations have the most potential to contribute to adaptive changes therefore likely
501 depends on both the shape of the corresponding fitness function and the respective distributions
502 of mutational effects (DME) for promoter and coding mutations. If the latter generally have
503 larger effects on molecular phenotypes like protein abundance and activity, they would be the
504 main contributor to adaptation when large changes are needed.
505
506 While the lack of strong mutational effects in the *FCY1* promoter could mean that this gene's
507 expression level is especially robust to mutations, multiple observations suggest that it is not
508 an exception in that regard. The substantial number of weak but significant expression changes
509 we report is first largely consistent with previous work. Large-scale mutagenesis of promoters
510 and regulatory elements in yeast [26,27] and mammals [30–33], have indeed shown that single
511 mutations at most positions have weak to moderate impact. The *FCY1* promoter is in addition
512 a typical yeast promoter, being TATA-less like the majority of such regions [46,60] and featuring
513 the most common poly-T/poly-A configuration [52]. Although bidirectional, it drives much
514 higher expression of *FCY1* than of the opposite gene and hundreds of such promoters have
515 been identified in the *S. cerevisiae* genome [61]. As such, the distribution of mutational effects
516 we report is likely generally representative of yeast promoters. Second, our work mirrors
517 previous accounts associating weaker functional impacts with changes within promoter
518 regions. A study of fluorescent reporter proteins for instance reported many more coding
519 sequence than cis-regulatory changes among fluorescence-decreasing mutations [62]. In addition,
520 a recent compilation of over 14 000 resistance-conferring mutations in 94 fungi uncovered only
521 34 promoter mutations, all of which involve insertions of tens of nucleotides [63].

More importantly, the discrepancy between the respective DME of promoter and coding sequence changes in *FCY1* is consistent with the general picture which emerged from earlier DMS studies of protein-coding sequences and regulatory regions [64]. While reported DMEs for proteins are often strikingly bimodal – with the abundance of strongly deleterious mutations sometimes comparable to that of WT-like mutants –, those of cis-regulatory elements tend to be unimodal and centered around the WT or to display a minor mode of a small number of strongly deleterious effects [64]. More recent mutagenesis studies have continued to frequently report distributions of effects which are consistent with this assertion [25,65–70], sometimes even showing that most mutations in a protein sequence are either highly deleterious or strongly beneficial [7].

This multimodality of coding sequence DMEs likely emerges from nonlinear relationships between the biophysical properties of proteins and fitness-relevant traits [64,71]. This can for instance be the case of protein stability, which dictates the fraction of correctly folded – and thus functional – proteins through a sigmoid function [70,72–74]. Due to this relationship, most mutations affecting stability would either have little effect or drastically decrease activity, for a WT protein which stably folds and is thus at the peak of the corresponding curve. This phenomenon likely underlies the high availability of 5-FC resistance through LOF mutations in the *FCY1* gene. Among resistance-conferring single-nucleotide substitutions in *FCY1*, only 16% (67/408) introduce a premature stop codon, while the predicted impact of mutations on protein stability is strongly correlated with fitness effects and resistant variants with reduced stability have been characterized experimentally [6,75]. Although similar nonlinearities are expected in the binding of regulators to DNA [64,76], almost entirely abolishing gene expression may be comparatively difficult to achieve. This is illustrated both by spurious transcription of the yeast genome [77] and by the fact that exogenous or random DNA introduced in yeast is generally transcribed [78,79]. As such, the unequal contributions of promoter and coding changes to 5-FC adaptation likely arise from fundamental differences in the effects of the two types of mutations. Although it is possible that this only applies when adaptation involves LOF, previous studies have shown that mutations in promoters can also be insufficiently adaptive in gain of function contexts. For instance, while increasing the expression of antifungal target *ERG11* through gene copy number variation (CNV) can confer resistance in fungi [80,81], no equally adaptive single promoter mutation has been reported [63]. Similarly, whereas CNVs of the sulfate transporter *SUL1* consistently arise in yeast evolved in sulfate-limited conditions [12,13], single mutations in the corresponding promoter have been shown to provide much less adaptive benefit [14].

Overall, it thus appears plausible that adaptation is generally more accessible through mutations in protein-coding sequences than in their cis-regulatory regions, at least when large changes to molecular traits like protein abundance and activity are needed (**Fig. 6a**). This may however vary across genes and adaptive phenotypes, since the availability of beneficial fitness effects through mutations in promoter and coding sequences depends on their respective DME and on the shape of the fitness function. A higher curvature around WT expression / activity levels could for instance make adaptation accessible through promoter as well as coding mutations, even if the former had smaller effects as in the *FCY1* model system (**Fig. 6b**). Other shapes of

fitness functions may in contrast restrict adaptation to small variations of molecular traits, potentially favoring cis-regulatory mutations (**Fig. 6c**). Additional studies are hence needed to fully map which contexts may favor one type of change over the other. With the current work, we have shown that combining the DMS of genes and cis-regulatory regions with the characterization of the corresponding fitness functions is a powerful approach to do so.

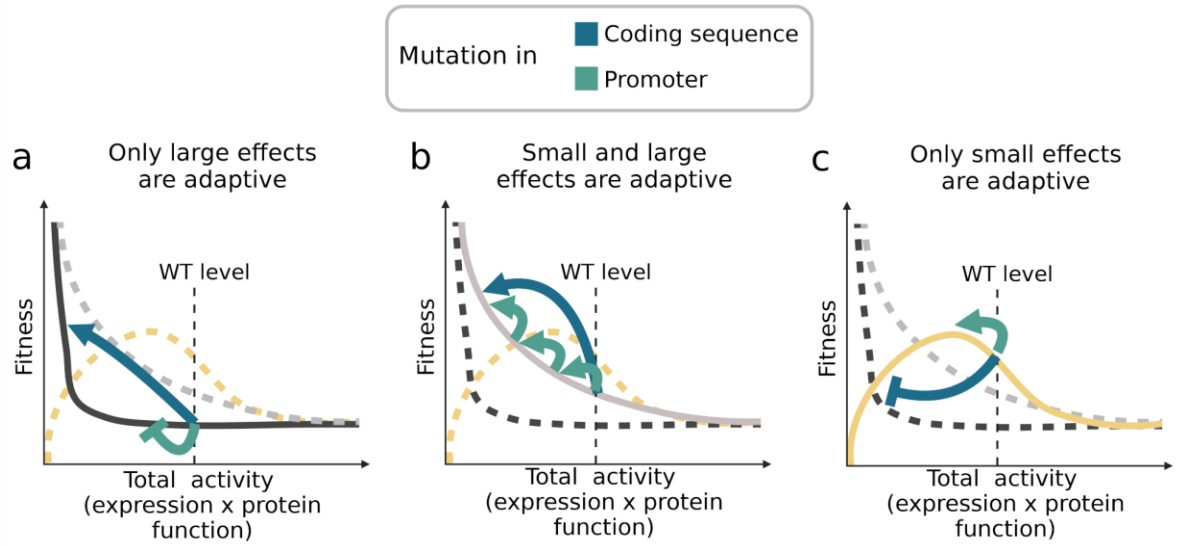


Fig. 6: Different shapes of fitness function may favor mutational effects of different magnitudes. Assuming that mutations in coding sequences have larger effects on molecular traits than changes within promoters, as is the case for *FCY1*, this can modulate whether adaptation is accessible through one type of mutations or the other. Although we only measured *FCY1* expression, we generalize expression-fitness relationships to total enzymatic activity. For each of the three scenarios, the corresponding fitness curve is shown as a solid line. **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/enzcymq.

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 [82] 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 in **Supplementary Table 4**. When specified, selective media was prepared by adding

one or multiple of the following agents: geneticin [200 µg/mL] (G418; Bioshop Canada), nourseothricin [100 µg/mL] (NAT; Jena Bioscience) and hygromycin B [250 µg/mL] (HYG; Bioshop Canada). All Sanger sequencing was performed at the sequencing platform of Centre de recherche du CHU de Québec-Université Laval (Université Laval, Québec, Canada).

Yeast transformation

Whenever inserts which included a selection marker had to be inserted in the yeast genome, a standard lithium acetate yeast transformation protocol was used [83]. After selection of transformants on the appropriate selective media, four to eight clones were selected for validation PCR, depending on the strain. Unless stated otherwise, colony PCR was performed on each side of the integration site to confirm proper integration, using Taq polymerase (Bioshop Canada). All genomic inserts were amplified using the KAPA HiFi HotStart DNA polymerase (Roche Molecular Systems).

This approach was for instance used to delete each fragment F1 to F4 of the intergenic region between *JID1* (*YPR061C*) and *FCY1* (*YPR062W*), as well as the full intergenic segment and the complete *FCY1* region (including the corresponding gene), by the insertion of a *natNT2* cassette [84]. This resulted in a set of six strains harboring a different deletion of the *FCY1* region (SA0127, SA0128, SA0129, SA0130, SA0131, SA0132). Similar deletions of regions F3 and F4 were also performed in the 42bp-mEGFP construct obtained by CRISPR-Cas9 yeast transformation (see below), generating fluorescent strains SA0246, SA0247 and SA0254

CRISPR-Cas9 yeast transformation

When sequences needed to be inserted in the yeast genome without any flanking selection marker, a CRISPR-Cas9 approach was used. This required the prior generation of a strain harboring a *natNT2* cassette at the chosen insertion site, which was done through standard yeast transformation as described above. Competent cells from this genetic background could then be cotransformed with the pCAS-NAT plasmid (**Supplementary Table 3**) – encoding a guide RNA engineered to target the Cas9 endonuclease to the *natNT2* cassette [39] – and the desired repair DNA, amplified with KAPA HiFi HotStart DNA polymerase (Roche Molecular Systems). A previously published CRISPR-Cas9 transformation protocol was followed [85]. Briefly, competent cells were prepared as described, and transformed with unpurified PCR amplification of the insert, DNA carrier [10 mg/ml], pCAS-NAT and PLATE solution (polyethylene glycol 2000 [40%], lithium acetate [0.1 M], Tris-EDTA [10 mM Tris-Cl, 1 mM EDTA] buffer [pH=8.0]). After mixing, cells were incubated for 30 minutes at 30°C, heat shocked for 15 minutes at 42°C, resuspended in YPD media and incubated for 4-6 hours at 30°C. Transformation reactions were plated on YPD media containing G418, to select for uptake of the pCas plasmid, and incubated at 30°C for 48 hours. Eight colonies from each transformation were then grown overnight (30°C with shaking) in liquid YPD and streaked on solid YPD. After 48 hours of incubation at 30°C, one isolated clone from each streaking was grown overnight in YPD (30°C with shaking) and spotted on both YPD and YPD+G418 solid media, to confirm the loss of the pCAS plasmid. Because the AKD1123 parental background is resistant to hygromycin, HYG was added to the YPD throughout. Correct integration was then validated by colony PCR and Sanger sequencing.

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

Growth assays

Unless stated otherwise, the following protocol was used each time the growth of *FCY1* strains was assessed in media containing 5-FC (Alfa Aesar) or cytosine (Alfa Aesar). Each yeast strain of interest was first inoculated in three replicates in SC [pH=6] liquid media – potentially with HYG, NAT and/or G418, depending on each strain's resistance marker(s) – and grown overnight at 30°C with shaking. After ~16 hours of incubation, a sample (100 µl) of each culture was centrifuged and resuspended in the same volume of sterile water, before being aliquoted and diluted again in sterile water to obtain up to 250 µl of cells at 1 optical density (OD) unit/ml. Growth assays were then assembled in Greiner flat-bottom transparent 96-well plates as follows: 160 µl of media (SC complete for 5-FC or SC-Ura for cytosine); 20 µl of diluted cells [OD=1]; 20 µl of 5-FC or cytosine diluted to obtain the desired final concentration in a volume of 200 µl. Each plate was incubated for 48 hours at 30°C without shaking in a Tecan Infinite M Nano (Tecan Life Sciences) and OD_{595} was measured at 15 minutes intervals.

To obtain exponential growth rates from the resulting growth curves, the complete time-series of OD_{595} was first log-transformed, after adjustment of values to a volume of 1 ml. Then, the slope between each consecutive OD_{595} value was computed, scaled according to the time interval of 0.25 h separating each measurement. For each unique growth curve, the median of the ten largest slopes was used as growth rate for the corresponding culture. When converting the growth rate of any strain i into selection coefficient s_i, the following formula was used [55]:

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

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

Construction of the mutant libraries

Amplification of mutated F3 and F4 promoter fragments

All single-nucleotide substitutions, insertions and deletions in the F3 and F4 regions were ordered as six separate oPools Oligo Pools (Integrated DNA Technologies Inc.) containing sequences of 117 bp, 118 bp and 116 bp, respectively. Within the different pools, each oligonucleotide is a variant of the full-length F3 or F4 region (77 bp) with flanking 20 bp homology regions at both ends. The latter sequences are respectively the 20 bp immediately upstream or downstream of the corresponding fragment in the reference genome of *S. cerevisiae* S288C, as retrieved from the Saccharomyces Genome Database [86]. For substitutions, each position of each fragment was iteratively degenerated into all four possible nucleotides by specifying a N base, while for insertions, the same mixed base was inserted in 3' of each position in the reference F3 and F4 sequences. When such a N is encountered, an equal ratio of each of the four standard nucleotides (A, T, G, C) is dispensed during synthesis. For the two pools of deletions, single-nucleotide removal was performed iteratively at each position of F3 and F4 and duplicates (occurring whenever two or more consecutive nucleotides are identical in the reference) were removed to obtain non-redundant sets. All sequences ordered as part of the six oPools are listed in **Supplementary Table 5**.

Upon reception, each oPool was spun down and resuspended in fresh PCR-grade water, to obtain a final concentration of 50 µM. Aliquots of 100 µl at 0.05 µM were then obtained by serial dilution from 2 µl of the initial resuspension. From these aliquots, PCR amplification of each oPool was performed in four replicates using KAPA HiFi HotStart DNA polymerase (Roche Molecular Systems) and the following

691 cycle: 98°C for 3 min; {98°C for 30 s; 60°C for 15 s; 72°C for 10 s} x 22; 72°C for 20 s; Hold at 10°C.
692 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**

693 In each case, the diluted oPool resuspension was not added to the reaction mix and was instead added
694 individually in each reaction tube. In addition, one control without the oPool dilution was performed
695 for each primer combination. Both primer pairs were designed to extend the 20 bp homology regions
696 included in the oPool, such that mutated F3 and F4 featured 60 bp homology arms upstream and
697 downstream of the corresponding genomic region. The two pairs were CLOP299-C3 / CLOP299-C4
698 and CLOP299-D3 / CLOP299-D4, respectively to amplify F3 and F4 oPools.
699
700 **Genomic insertion of mutated fragments in *S. cerevisiae***
701 To insert promoter mutations upstream of *FCY1*, competent yeast cells harboring a *natNT2* deletion of
702 the F3 or F4 region (strains SA0129 and SA0130) were transformed with the corresponding PCR-
703 amplified oPools. For each mutation category and each region, four replicate transformations were
704 performed from the corresponding four PCR reactions described above – for a total of 24
705 transformations. The protocol described by Ryan *et al.* (2016) [85] was followed (see *CRISPR-Cas9 yeast*
706 *transformation*, above). After incubation, G418-positive colonies from each agar plate were collected
707 separately. Liquid YPD (5 ml) was added to the plate, cells were resuspended by scraping with a glass
708 rake and the suspension was transferred to a 15 ml Falcon tube. From this suspension, glycerol stocks
709 (687 µl cells mixed with 313 µl sterile [80%] glycerol) were made for each transformation and stored
710 at -80°C. Two 500 µl aliquots from each tube were spun down and stored at -20°C after aspiration of
711 the supernatant. From these samples, mutant libraries were validated by PE150 sequencing on an
712 Illumina MiSeq instrument (Genomics Analysis Platform [PAG] of Institut de Biologie Intégrative et
713 des Systèmes, Université Laval, Québec). This approach was also employed to insert the same set of
714 promoter mutations upstream of the 42bp-mEGFP construct, using strains SA0246 and SA0254. In this
715 case, three replicate transformations were instead performed per amplified oPool.
716
717 **Selection of coding sequence mutations**
718 Fifteen codon positions of the *FCY1* gene, for which all possible states had previously been generated
719 by DMS [6], were selected to obtain a diversity similar to that of the complete mutational scanning of the
720 F3-F4 region (15 positions x 64 codons = 960 mutants). Among these positions, six were chosen for
721 their wide range of variant effects and nine others were selected randomly, to fully cover the previously
722 reported range of mutational effects. The 15 corresponding pools of mutants were each grown for ~24
723 hours in 4 ml SC complete media (30°C with shaking), from 25 µl of glycerol stock. Optical densities

were measured and all cultures were mixed in equal amounts to make four replicate 1 ml glycerol stocks (687 µl mixed cultures and 313 µl sterile [80%] glycerol) of this pool of Fcy1p mutants, which were stored at -80°C.

Pooling of mutants

The different sets of promoter mutations (substitutions, insertions and deletions, separately for F3 and F4) were grown separately in 4 ml SC complete media, from 50-100 µl of the corresponding glycerol stock. After ~16 hours of incubation at 30°C with shaking, OD_{595} was measured for all cultures. Subpools were then mixed to obtain equal proportions of all single mutants, according to their OD_{595} and how many unique genotypes they each contained.

For mutations in the *FCY1* background, the four replicate pools of coding sequence mutations were also grown from 75 µl of glycerol stock. All subpools were then combined as described, according to their replicate number (i.e. replicate 1 for F3 substitutions was mixed with replicate 1 of all other subpools). Each subpool of promoter mutations was thus included in only two final pools: one with added coding sequence mutations and the other without. For mutations in the 42bp-mEGFP background, all subpools of promoter variants were combined as described without the addition of other mutants, except that there were only three replicates of each.

Bulk competition experiment in 5-FC media

Each mutant pool was grown in two replicates in 4 ml SC complete media (30°C, 200 rpm) for ~20 hours, from 111 µl or 65 µl of the corresponding glycerol stock. These inoculum volumes were chosen to have at least 2000x cells for each mutant, respectively for pools with and without added coding sequence mutations. The competition experiment was initiated by diluting each preculture at final OD_{595} of 0.01 in 25 ml of SC complete media (with and without 5-FC [12 µM]) in 50 ml Falcon tubes. T0 samples were also prepared by centrifuging two microtubes containing 5 OD_{595} units for each preculture (850 g for 5 minutes; 4°C). After aspiration of supernatant, cell pellets were stored at -20°C. Newly inoculated cultures were incubated (30°C with shaking) in the dark until OD_{595} reached ~1.0. At this first timepoint T1, each culture was diluted in at final OD_{595} of 0.01 in 25 ml of fresh media. Additional cell pellets were prepared from the remaining cultures. The full volume was first centrifuged at 4°C (1000 g; 5 minutes). Cell pellets were resuspended in ~5 ml of supernatants and two samples of ~5 OD_{595} units were prepared as previously. This process was repeated at timepoint T2 when the new cultures reached OD_{595} ~1.0 and final T3 samples were prepared when the newly inoculated T2-T3 cultures reached the same OD_{595}. The timings of each timepoint and the corresponding OD_{595} values for all mutant pools are listed in **Supplementary Table 6**.

Sort-seq experiment

One final pool of F3-F4 mutants in the 42bp-mEGFP background was grown overnight from 100 µl glycerol stock in 5 ml of SC complete [pH = 6.0] media, at 30°C with shaking. After measurement of its OD_{595}, the culture was diluted to 0.15 OD/ml in 20 ml of fresh SC complete [pH = 6.0] and incubated for 5 hours at 30°C with shaking, to reach exponential phase (~0.5 OD/mL). From this, 4.5 ml were transferred to a 5 ml polystyrene round-bottom Falcon tube. The remaining 15.5 ml were centrifuged to collect cells, so that genotype frequencies before sorting could also be measured by sequencing.

A FACSMelody cell-sorter (BD Biosciences) was used to sort yeast cells based on their fluorescence intensity. Five bins were defined along the distribution of log_{10} fluorescence intensities, numbered 0 to 4 from the lowest to the highest fluorescence levels. More cells were collected from the most populated

771 parts of the distribution (50 000 cells each from bins 2 and 3, versus 10 000 cells each from bins 0,1
772 and 4). Each bin was sorted twice, resulting in two technical replicates, in a 96 well plate. The resulting
773 fractions were each used to inoculate 3 ml of SC complete [pH = 6.0] media in a 24 deep-well plate and
774 grown for 24h at 30°C with shaking (400 rpm). Each culture was spun down at 3000 rpm (1000 g) for
775 5 min and all pellets were washed with sterile water and transferred to a 1.5 ml microtube. All
776 microtubes were centrifuged at 15 000 rpm (21 000 g) for 1 min, supernatants were removed and pellets
777 were frozen at -80°C. This process was repeated two times for the two other biological replicates, such
778 that six replicate samples were obtained for each bin.
779

780 **DNA library preparation and sequencing**

781 The same approach was used for all sequencing libraries, for both the bulk competition and sort-seq
782 experiments. Genomic DNA was first extracted from each cell pellet using a standard phenol-
783 chloroform extraction protocol [87], including RNAse treatment. For the resequencing of bulk competition
784 samples in PE300, an additional RNAse treatment (10 minutes, 37°C) was performed on previously
785 purified DNA samples, followed by precipitation with sodium acetate and cold ethanol. A first PCR
786 was then performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs) to amplify the
787 genotyping region chosen for each type of samples, which included the F3-F4 promoter regions and
788 flanking sequences of varying lengths. This PCR step added Illumina primer-binding sites at both ends,
789 separated from the fragment of interest by small numbers of N bases (between 2 and 8; 10 total per PCR
790 amplicon). For the initial (PE150) sequencing of bulk competition samples, primers CLOP311-C6 to
791 CLOP314-H9 were used (as ordered in **Supplementary Table 2**). Two reactions were made from each
792 DNA extraction from cultures with added coding sequence mutants. One was specific to the F3-F4 pool
793 while the other targeted only the added *FCY1* variants. This was possible because the pool of coding
794 sequence mutations was generated from a codon-optimized sequence, harboring many nucleotide
795 changes from the WT [6]. For the PE300 resequencing of the same samples, only F3-F4 PCR were
796 performed, using oligos CLOP328-A6 to CLOP328-H7. Finally, for the sequencing of sort-seq bins,
797 also in PE300, oligos CLOP328-A6 to CLOP341-A10 were employed. In this case, 7 N bases were
798 added per amplicon.
799

800 In all cases, amplicons from this first PCR step were purified with AMPure XP magnetic beads
801 (Beckman Coulter Life Sciences) and diluted to 0.01 ng/µl. These dilutions were used as template for
802 a second set of PCR reactions using Q5 polymerase, which added Nextera barcodes and Illumina
803 adapters i5 and i7. A unique combination of i5 and i7 indexes was used for each DNA sample. The final
804 PCR products were also purified with AMPure XP beads, before quantification using a NanoDrop
805 spectrophotometer (Thermo Fisher Scientific). Amplicons from the same sequencing round (PE150
806 competition, PE300 competition or PE300 sort-seq) were mixed in equal proportions based on the
807 NanoDrop quantification and the resulting pool was sent for sequencing. PE150 sequencing was
808 performed on a NovaSeq Illumina instrument (sequencing platform of Centre de recherche du CHU de
809 Québec-Université Laval, Québec), while PE300 sequencing was performed on an AVITI instrument
810 from Element Biosciences (Genomics Analysis Platform [PAG] of Institut de Biologie Intégrative et
811 des Systèmes, Université Laval, Québec).
812

813 **Analysis of sequencing reads**

814 Following each sequencing run, reads were received already demultiplexed, according to the sample id
815 associated with each combination of i5 and i7 adaptors used. Read quality was first assessed using
816 FastQC (v 0.12.1) and MultiQC (v 1.12) [88]. Paired-end reads were next merged using PANDAseq (v
817 2.11) [89], with the '-k' parameter set to 4. The N positions which had been added at both ends during the

first PCR step of library construction were trimmed using command "fastx_filter" of VSEARCH (v 2.28.1) [90], according to a table listing the number of N added for each sample. The first and last 20 bp of the resulting sequences, which were the binding sites for the oligonucleotide primers used in the first PCR step, were also trimmed using the same command. Following these trimming steps, all identical sequences were aggregated with the "derep_fulllength" command of VSEARCH. Each unique sequence was aligned to the corresponding reference (depending on the type of sample and the length of the sequenced region) using the needle aligner from the EMBOSS suite (v 6.5.7) [91]. For promoter (F3-F4) libraries, the opening and extension of gaps was equally penalized (4,4), while penalties of 50 and 0.5 were respectively used for opening and extension for *FCY1* coding sequence libraries [6]. Custom Python functions were then used to call variants from each alignment. In addition to WT reads, only sequences harboring single-nucleotide mutations (F3-F4 libraries) or single-codon substitutions (*FCY1* libraries) were kept for downstream analysis. These variants were further filtered by length, to remove those that differed too much from the expected length of the PCR fragment. For bulk competition samples, read counts of each single mutant were converted to WT-relative frequencies. These read counts were instead used as is in the downstream analysis of sort-seq samples.

Estimation of fitness effects

Calculation of selection coefficients

For each bulk competition culture, time series of WT-relative genotype frequencies were reconstituted from variant frequencies obtained at the four timepoints (T0 to T3). The number of generations associated with each timepoint was obtained from the corresponding durations in hours, which varied between cultures, and the doubling time of the WT strain in the same conditions (SC media with or without 5-FC [12 µM]), as estimated from growth curve experiments. For each single-mutant genotype, the selection coefficient s was obtained by fitting a linear model using the number of generations and the log-transformed WT-relative 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) [92] 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.

865 **Removal of outliers and aggregation of replicates**
866 In 5-FC samples, mutations sometimes appeared beneficial in only one pair of cultures (coming from
867 the same transformation replicate), likely due to secondary resistance-conferring mutations (See
868 Supplementary Note). Outlier removal was thus performed, separately for two types of libraries
869 (promoter or coding sequence). The promoter (F3-F4) libraries were further separated depending on
870 whether coding-sequence mutants had been added in the corresponding cultures. Within each group of
871 samples, the median absolute deviation (MAD) of selection coefficients was first computed for each
872 mutant. This was then used to compute robust z-scores for each replicate measurement within each
873 genotype. The resulting distribution of robust z-scores calculated across all mutant genotypes was
874 compared and its 95th percentile was defined as the threshold to identify outlier replicates. The filtering
875 of outliers was however restricted to a subset of mutants which displayed a higher variance, to ensure
876 that small absolute variations between replicates were not identified as outliers. This was assessed from
877 each genotype's standard deviation, with the 95th percentile of the overall distribution used as threshold.
878 For F3-F4 sequencing libraries obtained from cultures with added coding sequence mutations, the 85th
879 percentile was instead used.
880
881 To obtain final estimates of the fitness effects of promoter mutations, only F3-F4 libraries without added
882 coding sequence mutants were taken into account. Four of the eight replicate cultures, corresponding to
883 two of the four transformation replicates, were additionally ignored. In these four samples, selection
884 coefficients displayed a temporal shift towards negative values which appeared consistent with some
885 undetected resistance-conferring mutations being genotyped as WT (see Supplementary Note). For each
886 of the 1020 assayed promoter mutations, the corresponding fitness effect was estimated as the median
887 of the selection coefficients observed across the four remaining replicates, which were considered
888 unbiased. The statistical significance of selection coefficients was assessed from the same four replicate
889 measurements for each mutation, by comparing their mean to 0 using the SciPy (v 1.15.2) [93]
890 implementation of the one-sample t-test, followed by Benjamini-Hochberg correction controlling the
891 false discovery rate at 5%. For the coding sequence mutations added as controls, fitness effects were
892 instead estimated as the median of the eight replicate measurements, without ignoring any of the
893 corresponding samples.
894
895 **Comparison with published effects of amino acid substitutions in Fcy1p**
896 The selection coefficients of added coding sequence mutations were first aggregated into grand
897 medians, according to the corresponding amino acid substitutions. They were then compared with
898 previously published variant effect scores obtained in the same conditions (12 μM 5-FC) [6]. The
899 agreement between the two experiments was further assessed by classifying variants as resistant and
900 susceptible in both cases. To this end, Gaussian mixture models with two components and a "tied"
901 covariance matrix were fitted on each of the two datasets, using the scikit-learn (v 1.6.1) [94]
902 implementation.
903
904 For the comparison of promoter and coding sequence mutations, the full dataset from [6] was used. All
905 single-nucleotide substitutions in the reference sequence of the *FCY1* gene, obtained from the
906 Saccharomyces Genome Database [86], were first performed computationally. The corresponding variant
907 effects were retrieved from the Després *et al.* (2022) data [6], according to the resulting amino acid
908 substitutions. Of 1431 possible single-nucleotide substitutions in *FCY1*, only 20 were not covered,
909 mostly in the first and last codons. The Gaussian mixture model previously fitted on the full set of amino
910 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)

955 substitutions between positions -112 to -109 (n=2); 3) substitutions between positions -63 to -57 (n=2)
956 and 4) variants among the 25 highest scores (n=4).
957
958 The corresponding validation mutants were individually reconstructed using the CRISPR-Cas9 yeast
959 transformation protocol described above. For each genotype, this was performed in two genetic
960 backgrounds: fluorescent strain SA0247 (ΔF3-F4 with 42bp-mEGFP construct) and *FCY1* deletion
961 (ΔF3-F4) mutant SA0159 (see **Supplementary Table 1**). A fusion PCR product harboring the single
962 mutation of interest was used as donor DNA. This amplicon was generated by two consecutive rounds
963 of PCR. In the first round, two reactions were performed with WT genomic DNA as template. The first
964 one used a generic forward primer hybridizing upstream of F3 (CLOP326-A11) with a mutant-specific
965 reverse primer harboring the mutation of interest (**Supplementary Table 2**). In the second reaction, a
966 forward primer with the mutation of interest (**Supplementary Table 2**) was used alongside a generic
967 reverse primer binding either in the mEGFP sequence (CLOP326-G10; for background SA0247) or in
968 the *FCY1* coding sequence (CLOP326-E10; for background SA0159). The two mutant-specific
969 mutagenesis primers are complementary to each other, such that there is homology between the two
970 amplicons. The second PCR round used the two PCR products from the first round diluted 1/40 as
971 template and the same two generic primers used in the first pair of PCR (CLOP326-A11 with either
972 CLOP326-G10 or CLOP326-E10, depending on the mutagenized background), which allowed fusion
973 of both PCR parts. This approach thus allows to amplify the full F3-F4 promoter region (with a short
974 segment of *FCY1*) while introducing a single mutation within it and adding homology regions on both
975 ends for subsequent genomic insertion. It was additionally used to perform deletion of short segments
976 of the promoter (-133 to -127 and -125 to -118), similarly in both genetic backgrounds. After
977 transformation of each fusion PCR product, clones from each transformation were sent for Sanger
978 sequencing to validate proper genomic integration of the mutation. All 66 resulting strains, including
979 reconstructions of the WT promoter, are listed in **Supplementary Table 1**.

980 **Expression measurements by flow cytometry**

981 Measurement of promoter activity using fluorescent strains was performed with a Guava EasyCyte HT
982 cytometer (CytekBio). Each strain of interest was grown overnight in three replicates in SC complete
983 media at 30°C with shaking, along with similar replicate cultures of a reference strain with the WT
984 promoter (SA0247 or AKD1462) and of a non-fluorescent control (AKD1123 or AKD1429). The
985 overnight cultures were diluted at 0.15 OD_{595} in fresh SC-Trp (MSG) media and grown at 30°C with
986 shaking until ~0.5 OD_{595}. Exponentially growing cells were then diluted to 0.05 OD_{595} in fresh media
987 and single-cell fluorescence intensity measurements were performed with the Guava EasyCyte HT
988 system. For each replicate culture of each strain, 5000 events were acquired (488 nm excitation; 525
989 nm detection). The medians of the corresponding distributions of fluorescence intensities were used to
990 estimate relative expression levels.
991
992 Within each cytometry experiment, events were first filtered out according to forward scatter and side
993 scatter values, with the 99th percentile of each distribution used as thresholds. All fluorescence
994 intensities were then normalized by the corresponding forward scatter measurement. From these
995 normalized fluorescence intensities, a median was computed for each replicate culture assayed. These
996 medians were then normalized by the grand median of measurements made for the reference strain,
997 such that they were converted into WT-relative fluorescence intensities. The resulting values, and
998 especially their medians, were used as estimates of relative expression levels. Accordingly, three
999 replicate measurements of WT-relative fluorescence intensity were obtained for each tested strain, with
1000 a value of 1 being equal to the WT level.
1001

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) [92] 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

Construction of strains with distinct *FCY1* expression levels

Six yeast promoters covering a wide range of expression levels were used to construct two sets of strains, respectively from *FCY1* background AKD1123 and from fluorescent background SA0228. The six promoter cassettes, each including a G418 resistance marker, were PCR-amplified from the corresponding plasmids [84] (**Supplementary Table 3**) using oligonucleotides CLOP338-G6 and CLOP338-H6 (**Supplementary Table 2**). The two primers provided homology for genomic insertion immediately upstream of *FCY1* (or of its first 42 bp for the SA0228 background), between the gene and its native promoter. Yeast transformations were performed in the two genetic backgrounds as previously described, using each of the corresponding PCR products. Transformants were selected on YPD+HYG+G418 media. Correct integration of the promoter cassette was validated by PCR in four clones from each transformation, using oligonucleotides CLOP310-D1 and CLOP291-A3 (**Supplementary Table 2**). Two to three validated clones were obtained in most cases, such that

biological replicates could be used in downstream assays for the majority of genotypes. All 25 corresponding yeast strains are listed in **Supplementary Table 1**.

Measurements of growth rates and fluorescence intensities

The set of *FCY1* strains (background AKD1123) was used to measure growth rates in liquid media containing different concentrations of 5-FC (3 μM, 12 μM, 24 μM and 48 μM) or cytosine (47 μM, 63 μM, 84 μM and 105 μM), following the growth assays protocol described above. The set of fluorescent strains (background SA0228) was instead used in flow cytometry experiments to measure expression levels (see Expression measurements by flow cytometry). In both cases, all available biological replicates were used for each genotype. When less than three strains were available, three replicate cultures were nonetheless inoculated.

The growth assays in 5-FC media were performed in two separate experiments. More than three replicate measurements were thus available for two concentrations (12 μM and 3 μM). Growth rates from the first experiment were adjusted to have the same grand medians as those of the second one, by the addition of the difference between the two experiments' grand medians. This was performed separately for each concentration and allowed to combine the two sets of growth rates, in spite of potential systematic differences between them – for instance if the true concentration of 5-FC in the media varied. The cytometry measurements, which were done in parallel with the growth assays, were also performed in two separate experiments. In this case, potential systematic differences were controlled for by computing WT-relative fluorescence intensities within each experiment and only combining measurements after this step.

Fitting of expression-fitness functions

Fluorescence measurements associated with each promoter were combined with the corresponding growth rates. An exponential function was fitted for each of the eight conditions, using WT-relative fluorescence as the independent variable and growth rate as the dependent variable. This could model the sharp decline observed in 5-FC as well as the steep increase seen in cytosine conditions. The fitting of all expression-fitness curves was performed within a bootstrapping routine. For each promoter, two WT-relative fluorescence intensity measurements were first sampled with replacement from each cytometry experiment. Then, the four resulting values were resampled with replacement. Corresponding growth rates were sampled similarly. If a given condition had been tested in two experiments, two replicate growth rates were sampled with replacement from each experiment for each promoter, and the four measurements were afterwards resampled with replacement. This approach was employed to avoid favoring one experiment over the other, since the number of replicates sometimes varied between them. When only one experiment had been performed, replicate growth rates from this assay were instead resampled with replacement. Once relative fluorescence intensities and growth rates had been sampled, values obtained for the same promoter were paired randomly into data points. The exponential function was then fitted using the "optimize.curve_fit" implementation of the SciPy module (v 1.15.2) [93] module and evaluated over 250 equally spaced values covering the full range of WT-relative fluorescence intensities. After 5000 iterations, medians and corresponding 95% prediction intervals were computed for each of these 250 fluorescence levels, thus generating the expression-fitness function. This was repeated for each of the 5-FC and cytosine concentrations tested.

Validation of *FCY1* promoter variants fitness effects

The previously described validation mutants reconstructed in the SA0159 *FCY1* background (**Supplementary Table 1**) were grown in SC complete (MSG) media with 12 μM 5-FC and without,

according to the growth assays protocol described above. One genotype (AKD1417; C to G substitution at position -123) displayed both an elevated fitness in SC media and an outlier low growth rate in 5-FC conditions and was not included in downstream analyses. For all other mutants, growth rates measured in 12 µM 5-FC were converted into selection coefficients as described in the growth assays section, using growth rates measured in the same experiment for WT strain AKD1123. The expression-fitness function obtained in the same conditions and the bounds of its 95% prediction interval were also converted into selection coefficients using the same approach. In this case, the WT growth rate interpolated from the curve was instead used as comparison.

Assessment of 5-FC resistance arising from mutant library construction

To test whether mutations conferring 5-FC resistance were introduced during mutant library construction, the process described in *Construction of mutant libraries* was repeated, but using WT sequences of the F3 and F4 regions. Briefly, the same two oligonucleotide pairs (CLOP299-C3/CLOP299-C4 and CLOP299-D3/CLOP299-D4) were first used to amplify F3 and F4 from genomic DNA of the AKD1123 strain, using KAPA HiFi HotStart DNA polymerase (Roche Molecular Systems). Competent cells from strains SA0129 and SA0130 (prepared as described by [85]) were then transformed with the F3 and F4 inserts, respectively. Two biological replicates of each transformation were made. This was performed following the CRISPR-Cas9 approach described above for the pools of single mutations. After 48 hours of incubation at 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 ($s \sim \mathcal{N}(0, 0.005)$) 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 0.4 to 0.6, 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) [88], PANDAseq (v 2.11) [89], VSEARCH (v 2.28.1) [90] and EMBOSS suite (v 6.5.7) [91]. 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

CRL, AKD and SA designed the study. SA and AKD performed the experiments as well as the data analysis. SA, AKD and CRL interpreted the results. SA, AKD and CRL wrote the manuscript.

References

1. Wright, S. The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. *Proceedings of the XI International Congress of Genetics* **8**, 209–222 (1932).

2. Smith, J. M. Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).

3. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).

4. Aguilar-Rodríguez, J., Payne, J. L. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* **1**, 45 (2017).

5. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).

6. Després, P. C. *et al.* Asymmetrical dose responses shape the evolutionary trade-off between antifungal resistance and nutrient use. *Nat. Ecol. Evol.* **6**, 1501–1515 (2022).

7. Bédard, C. *et al.* Most azole resistance mutations in the Candida albicans drug target confer cross-resistance without intrinsic fitness cost. *Nat. Microbiol.* **9**, 3025–3040 (2024).

8. Rouleau, F. D. *et al.* Deep mutational scanning of Pneumocystis jirovecii dihydrofolate reductase reveals allosteric mechanism of resistance to an antifolate. *PLoS Genet.* **20**, e1011252 (2024).

1224 9. Jiang, L. *et al.* A Balance between Inhibitor Binding and Substrate Processing Confers
1225 Influenza Drug Resistance. *Journal of molecular biology* **428**, (2016).
1226 10. Flynn, J., Samant, N., Schneider-Nachum, G., Tenzin, T. & Bolon, D. N. A. Mutational
1227 fitness landscape and drug resistance. *Curr. Opin. Struct. Biol.* **78**, 102525 (2023).
1228 11. Estevam, G. O. *et al.* Mapping kinase domain resistance mechanisms for the MET receptor
1229 tyrosine kinase via deep mutational scanning. *Elife* **13**, (2025).
1230 12. Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to controlled
1231 nutrient-limited environments in yeast. *PLoS Genet.* **4**, e1000303 (2008).
1232 13. Payen, C. *et al.* The dynamics of diverse segmental amplifications in populations of
1233 Saccharomyces cerevisiae adapting to strong selection. *G3 (Bethesda)* **4**, 399–409 (2014).
1234 14. Rich, M. S. *et al.* Comprehensive Analysis of the SUL1 Promoter of Saccharomyces
1235 cerevisiae. *Genetics* **203**, 191–202 (2016).
1236 15. Ascencio, D. *et al.* Expression attenuation as a mechanism of robustness against gene
1237 duplication. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2014345118 (2021).
1238 16. Siddiq, M. A., Duveau, F. & Wittkopp, P. J. Plasticity and environment-specific relationships
1239 between gene expression and fitness in Saccharomyces cerevisiae. *Nat. Ecol. Evol.* **8**, 2184–2194
1240 (2024).
1241 17. Siu, L. K., Ho, P. L., Yuen, K. Y., Wong, S. S. & Chau, P. Y. Transferable hyperproduction
1242 of TEM-1 beta-lactamase in Shigella flexneri due to a point mutation in the pribnow box. *Antimicrob.*
1243 *Agents Chemother.* **41**, 468–470 (1997).
1244 18. Jay, A., Jordan, D. F., Gerstein, A. & Landry, C. R. The role of gene copy number variation
1245 in antimicrobial resistance in human fungal pathogens. *NPJ Antimicrob Resist* **3**, 1 (2025).
1246 19. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein.
1247 *Nature* **436**, 588–592 (2005).
1248 20. Jiang, L., Mishra, P., Hietpas, R. T., Zeldovich, K. B. & Bolon, D. N. A. Latent effects of
1249 Hsp90 mutants revealed at reduced expression levels. *PLoS Genet.* **9**, e1003600 (2013).
1250 21. Rest, J. S. *et al.* Nonlinear fitness consequences of variation in expression level of a
1251 eukaryotic gene. *Mol. Biol. Evol.* **30**, 448–456 (2013).
1252 22. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on
1253 Fitness. *Cell* **166**, 1282–1294.e18 (2016).
1254 23. Arita, Y. *et al.* A genome-scale yeast library with inducible expression of individual genes.
1255 *Mol. Syst. Biol.* **17**, e10207 (2021).
1256 24. Cisneros, A. F. *et al.* Epistasis between promoter activity and coding mutations shapes gene
1257 evolvability. *Sci. Adv.* **9**, eadd9109 (2023).
1258 25. Tsuru, S., Hatanaka, N. & Furusawa, C. Promoters constrain evolution of expression levels of
1259 essential genes in Escherichia coli. *Mol. Biol. Evol.* **41**, msae185 (2024).
1260 26. Metzger, B. P. H. *et al.* Contrasting Frequencies and Effects of cis- and trans-Regulatory

Mutations Affecting Gene Expression. *Mol. Biol. Evol.* **33**, 1131–1146 (2016).

27. Hodgins-Davis, A., Duveau, F., Walker, E. A. & Wittkopp, P. J. Empirical measures of mutational effects define neutral models of regulatory evolution in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* **116**, 21085–21093 (2019).

28. Vaishnav, E. D. *et al.* The evolution, evolvability and engineering of gene regulatory DNA. *Nature* **603**, 455–463 (2022).

29. Li, X. C., Fuqua, T., van Breugel, M. E. & Crocker, J. Mutational scans reveal differential evolvability of Drosophila promoters and enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **378**, 20220054 (2023).

30. Patwardhan, R. P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).

31. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).

32. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).

33. Kircher, M. *et al.* Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat. Commun.* **10**, 3583 (2019).

34. Duveau, F. *et al.* Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. *Elife* **7**, e37272 (2018).

35. Ringel, A. E. *et al.* Yeast Tdh3 (glyceraldehyde 3-phosphate dehydrogenase) is a Sir2-interacting factor that regulates transcriptional silencing and rDNA recombination. *PLoS Genet.* **9**, e1003871 (2013).

36. Kankipati, H. N., Rubio-Texeira, M., Castermans, D., Diallinas, G. & Thevelein, J. M. Sul1 and Sul2 sulfate transceptors signal to protein kinase A upon exit of sulfur starvation. *J. Biol. Chem.* **290**, 10430–10446 (2015).

37. Vermes, A., Guchelaar, H. J. & Dankert, J. Flucytosine: a review of its pharmacology, clinical indications, pharmacokinetics, toxicity and drug interactions. *J. Antimicrob. Chemother.* **46**, 171–179 (2000).

38. Lin, Z., Wu, W.-S., Liang, H., Woo, Y. & Li, W.-H. The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. *BMC Genomics* **11**, 581 (2010).

39. Lemieux, P., Bradley, D., Dubé, A. K., Dionne, U. & Landry, C. R. Dissection of the role of a Src homology 3 domain in the evolution of binding preference of paralogous proteins. *Genetics* **226**, iyad175 (2024).

40. Dionne, U. *et al.* Protein context shapes the specificity of SH3 domain-mediated interactions in vivo. *Nat. Commun.* **12**, 1597 (2021).

41. Matuszewski, S., Hildebrandt, M. E., Ghenu, A.-H., Jensen, J. D. & Bank, C. A statistical guide to the design of deep mutational scanning experiments. *Genetics* **204**, 77–87 (2016).
42. de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* **38**, 56–65 (2020).
43. Peterman, N. & Levine, E. Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics* **17**, 206 (2016).
44. Eisenberg, A. R. *et al.* Translation initiation site profiling reveals widespread synthesis of non-AUG-initiated protein isoforms in yeast. *Cell Syst.* **11**, 145–160.e5 (2020).
45. de Boer, C. G. & Hughes, T. R. YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res.* **40**, D169–79 (2012).
46. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
47. McMillan, J., Lu, Z., Rodriguez, J. S., Ahn, T.-H. & Lin, Z. YeasTSS: an integrative web database of yeast transcription start sites. *Database (Oxford)* **2019**, baz048 (2019).
48. Huang, Q., Szklarczyk, D., Wang, M., Simonovic, M. & von Mering, C. PaxDb 5.0: Curated protein quantification data suggests adaptive proteome changes in yeasts. *Mol. Cell. Proteomics* **22**, 100640 (2023).
49. Erb, I. & van Nimwegen, E. Transcription factor binding site positioning in yeast: proximal promoter motifs characterize TATA-less promoters. *PLoS One* **6**, e24279 (2011).
50. Pelechano, V., Wei, W. & Steinmetz, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127–131 (2013).
51. Struhl, K. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8419–8423 (1985).
52. de Boer, C. G. & Hughes, T. R. Poly-dA:dT tracts form an in vivo nucleosomal turnstile. *PLoS One* **9**, e110479 (2014).
53. Ichikawa, Y. *et al.* Sequence-directed nucleosome-depletion is sufficient to activate transcription from a yeast core promoter in vivo. *Biochem. Biophys. Res. Commun.* **476**, 57–62 (2016).
54. Jansen, A., van der Zande, E., Meert, W., Fink, G. R. & Verstrepen, K. J. Distal chromatin structure influences local nucleosome positions and gene expression. *Nucleic Acids Res.* **40**, 3870–3885 (2012).
55. Chevin, L.-M. On measuring selection in experimental evolution. *Biol. Lett.* **7**, 210–213 (2011).
56. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation: The locus of evolution. *Evolution* **61**, 995–1016 (2007).
57. Delma, F. Z. *et al.* Genetic mutations in Cryptococcus neoformans pyrimidine salvage pathway enzymes contribute to reduced susceptibility against 5-fluorocytosine. *NPJ Antimicrob.*

Resist. **2**, 24 (2024).
58. Kruglyak, L. *et al.* Insufficient evidence for non-neutrality of synonymous mutations. *Nature* **616**, E8–E9 (2023).
59. 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.
60. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
61. 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).
62. Gruber, J. D., Vogel, K., Kalay, G. & Wittkopp, P. J. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in Saccharomyces cerevisiae: frequency, effects, and dominance. *PLoS Genet.* **8**, e1002497 (2012).

63. Bédard, C. *et al.* FungAMR: a comprehensive database for investigating fungal mutations associated with antimicrobial resistance. *Nat. Microbiol.* 1–15 (2025) doi:10.1038/s41564-025-02084-7.
64. Kemble, H., Nghe, P. & Tenaillon, O. Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. *Evol. Appl.* **12**, 1721–1742 (2019).
65. Vanella, R. *et al.* Understanding activity-stability tradeoffs in biocatalysts by enzyme proximity sequencing. *Nat. Commun.* **15**, 1807 (2024).
66. Amorosi, C. J. *et al.* Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *Am. J. Hum. Genet.* **108**, 1735–1751 (2021).
67. Prywes, N. *et al.* A map of the rubisco biochemical landscape. *Nature* **638**, 823–828 (2025).
68. Grønbæk-Thygesen, M. *et al.* Deep mutational scanning reveals a correlation between degradation and toxicity of thousands of aspartoacylase variants. *Nat. Commun.* **15**, 4026 (2024).
69. Choudhury, A. *et al.* Deep mutational scanning reveals the molecular determinants of RNA polymerase-mediated adaptation and tradeoffs. *Nat. Commun.* **14**, 6319 (2023).
70. Beltran, A., Jiang, X. ’er, Shen, Y. & Lehner, B. Site-saturation mutagenesis of 500 human protein domains. *Nature* **637**, 885–894 (2025).
71. Otwinowski, J., McCandlish, D. M. & Plotkin, J. B. Inferring the shape of global epistasis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7550–E7558 (2018).
72. Wylie, C. S. & Shakhnovich, E. I. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9916–9921 (2011).
73. Serohijos, A. W. R. & Shakhnovich, E. I. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr. Opin. Struct. Biol.* **26**, 84–91 (2014).

74. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).
75. Després, P. C. *et al.* Compensatory mutations potentiate constructive neutral evolution by gene duplication. *Science* **385**, 770–775 (2024).
76. Lagator, M., Paixão, T., Barton, N. H., Bollback, J. P. & Guet, C. C. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *Elife* **6**, (2017).
77. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
78. Luthra, I. *et al.* Regulatory activity is the default DNA state in eukaryotes. *Nat. Struct. Mol. Biol.* **31**, 559–567 (2024).
79. Meneu, L. *et al.* Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus. *Science* **387**, eadm9466 (2025).
80. Cools, H. J., Hawkins, N. J. & Fraaije, B. A. Constraints on the evolution of azole resistance in plant pathogenic fungi. *Plant Pathol.* **62**, 36–42 (2013).
81. Todd, R. T. & Selmecki, A. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *Elife* **9**, e58349 (2020).
82. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
83. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
84. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
85. Ryan, O. W., Poddar, S. & Cate, J. H. D. CRISPR-Cas9 Genome Engineering in Saccharomyces cerevisiae Cells. *Cold Spring Harb. Protoc.* **2016**, db.prot086827 (2016).
86. Engel, S. R. *et al.* Saccharomyces Genome Database: advances in genome annotation, expanded biochemical pathways, and other key enhancements. *Genetics* **229**, iyae185 (2025).
87. Amberg, D. C., Burke, D. J. & Strathern, J. N. *Methods in Yeast Genetics 2005: A Cold Spring Harbor Laboratory Course Manual*. (Cold Spring Harbor Laboratory Press, New York, NY, 2005).
88. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
89. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
90. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
91. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European molecular biology open

software suite. *Trends Genet.* **16**, 276–277 (2000).

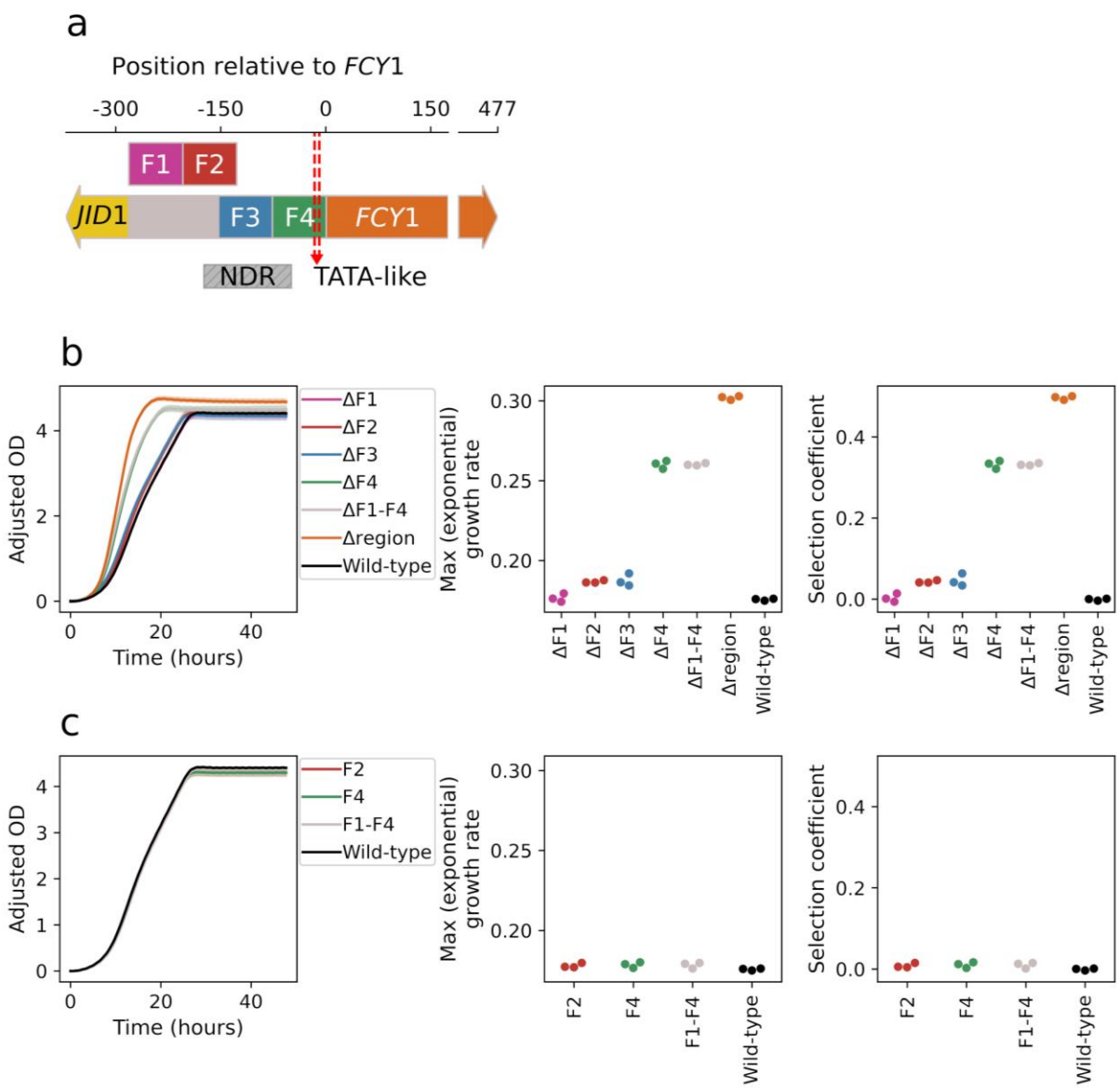
92. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. in *9th Python in Science Conference* (2010).

93. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

94. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

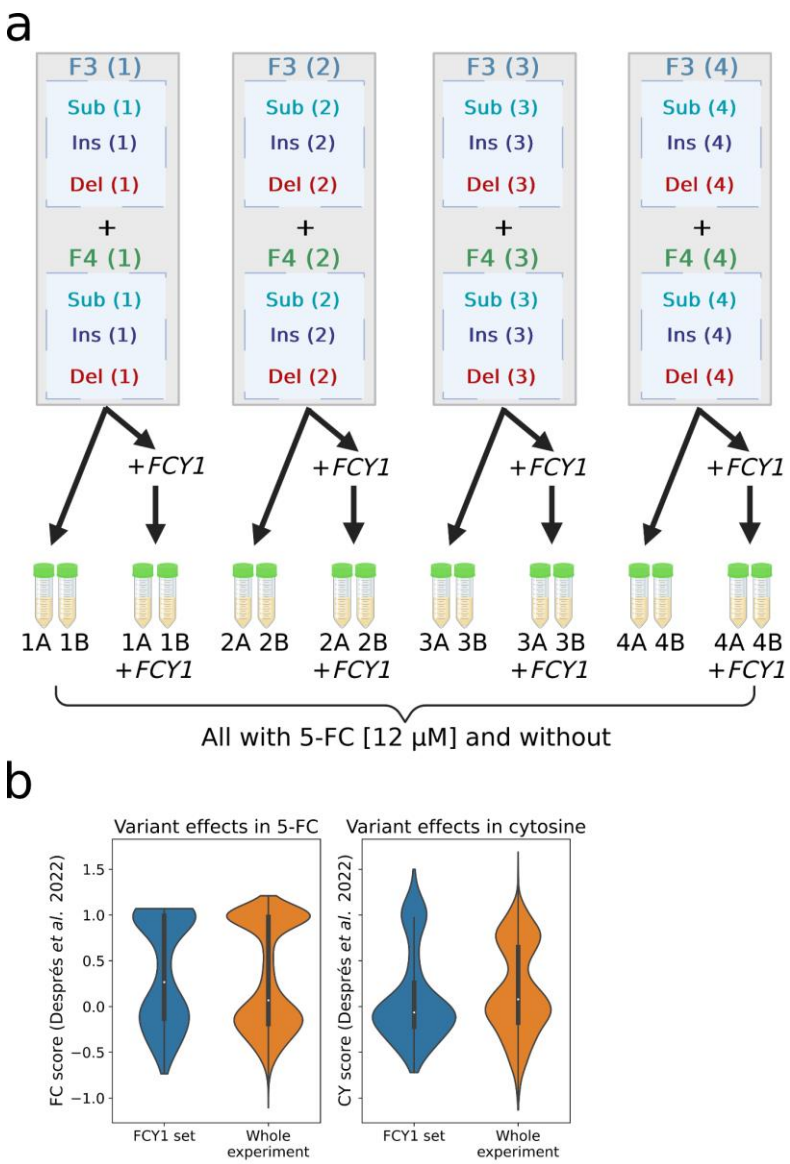
95. Chereji, R. V., Ramachandran, S., Bryson, T. D. & Henikoff, S. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol.* **19**, 19 (2018).

1433 **Extended Data Figures:**

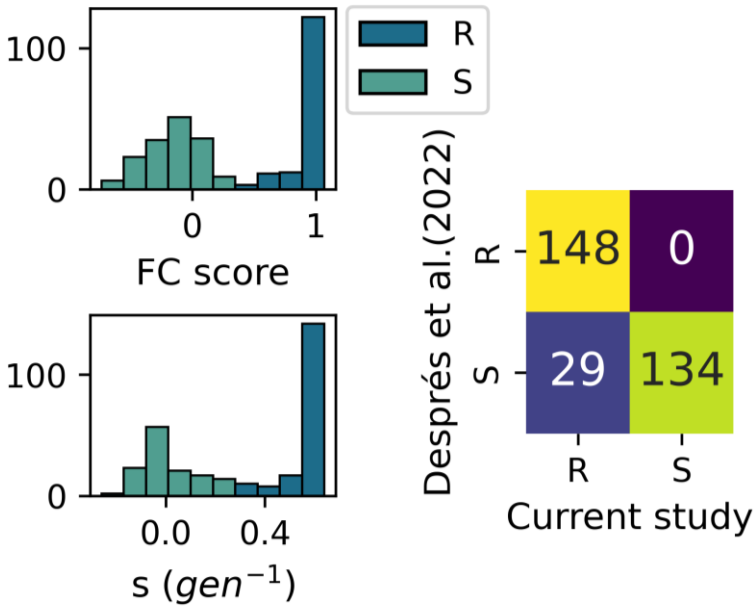


1434

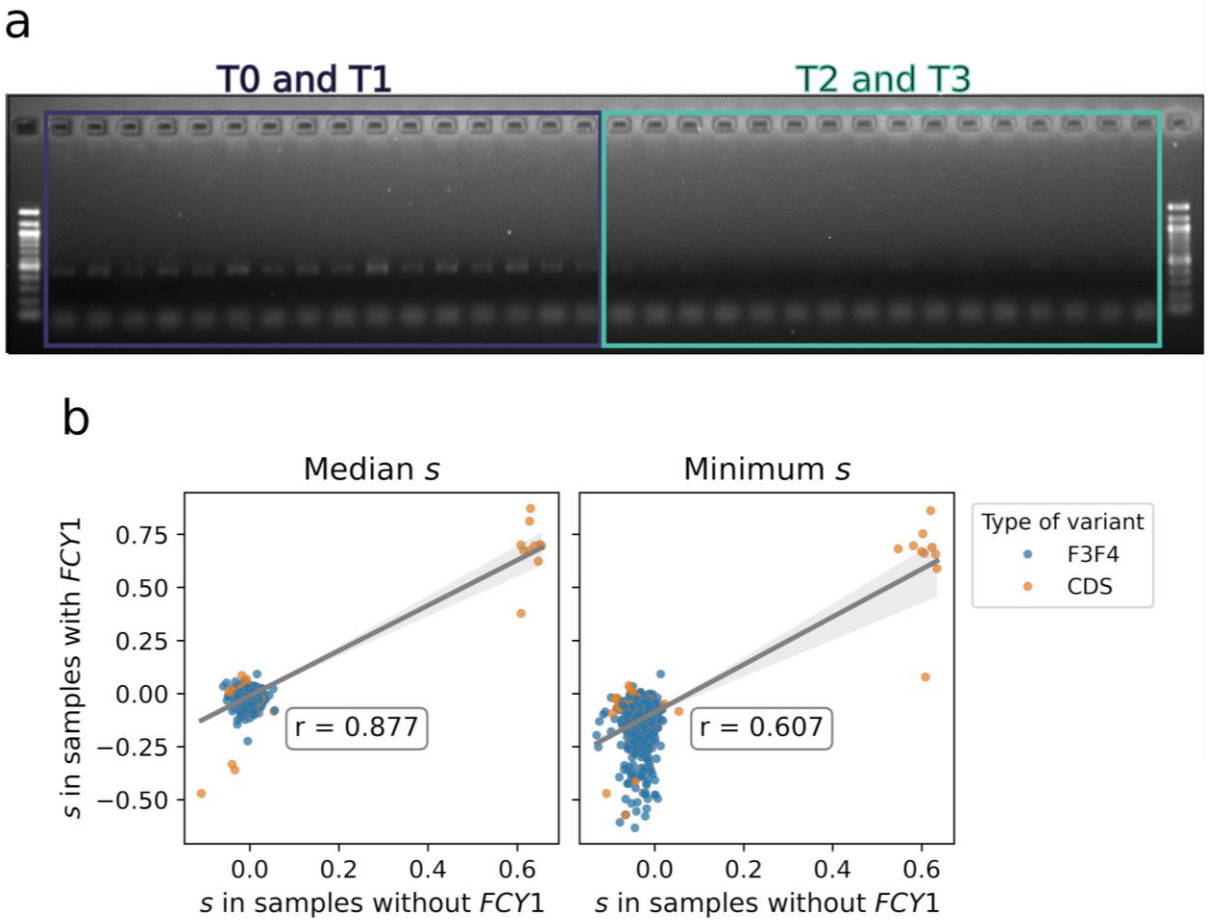
1435 **Extended Data Fig. 1**: **The deletion of the F4 promoter fragment is clearly adaptive in 5-FC**
1436 **conditions. a,** Schematics of the four 77 bp fragments into which the intergenic region between *JID1*
1437 (YPR061C) and *FCY1* (YPR062W) was divided. There is a 26 bp overlap between F2 and F3. The
1438 nucleosome-depleted region (NDR) of the corresponding bidirectional promoter [61,95] as well as the
1439 TATA-like element, involved in transcription initiation at *FCY1* [46,47], are shown. **b,** Fitness of deletion
1440 mutants in [12 μM] 5-FC. The ΔF1-F4 mutant refers to deletion of the complete intergenic region, while
1441 Δregion instead represents the simultaneous deletion of the intergenic region and the *FCY1* gene.
1442 Exponential growth rates and the corresponding selection coefficients (*Methods*) show that F4 is the
1443 only promoter fragment for which deletion confers a clear growth advantage (and thus 5-FC resistance).
1444 **c,** Fitness of reconstructed WT strains in [12 μM] 5-FC. Here, F2, F4 and F1-F4 refer to CRISPR
1445 reconstruction of the WT intergenic region from the corresponding deletion mutant (*Methods*). In all
1446 three cases, the WT 5-FC sensitivity is recovered. All growth measurements were made from three
1447 replicate cultures of each strain.



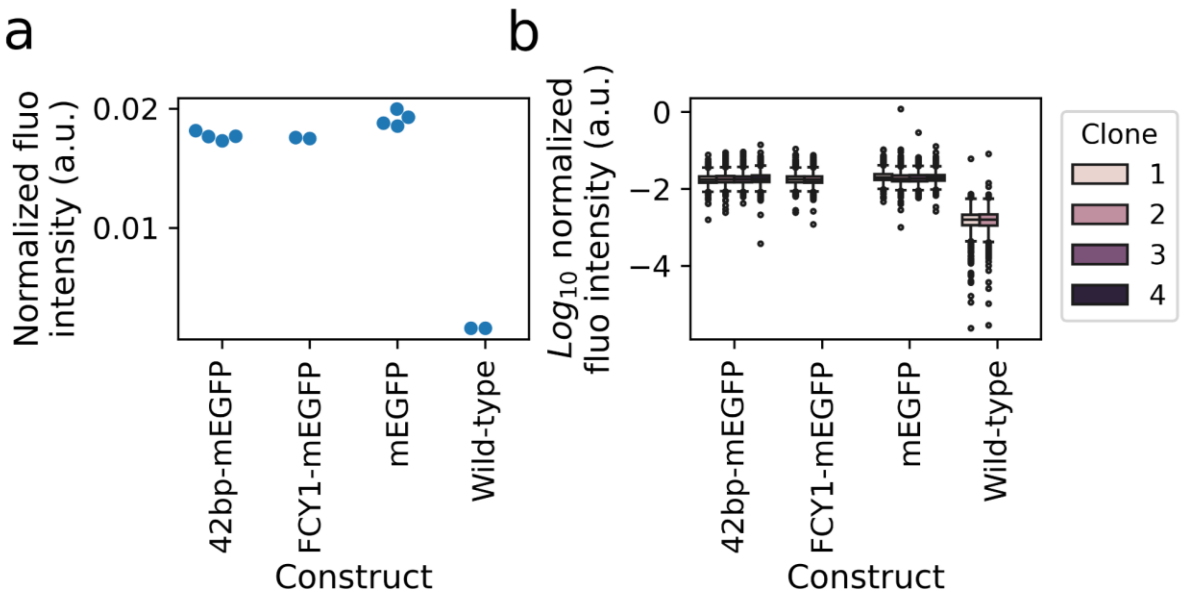
Extended Data Fig. 2: Complete overview of the bulk competition experiment. a, Full experimental design showing all 32 cultures and their relationships as transformation and/or culture replicates. Each pool of substitutions (Sub), insertions (Ins) or deletions (Del) was constructed four times and corresponding 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). 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 [6], both in 5-FC and cytosine (without uracil) conditions, where Fcy1 activity is respectively toxic and essential.



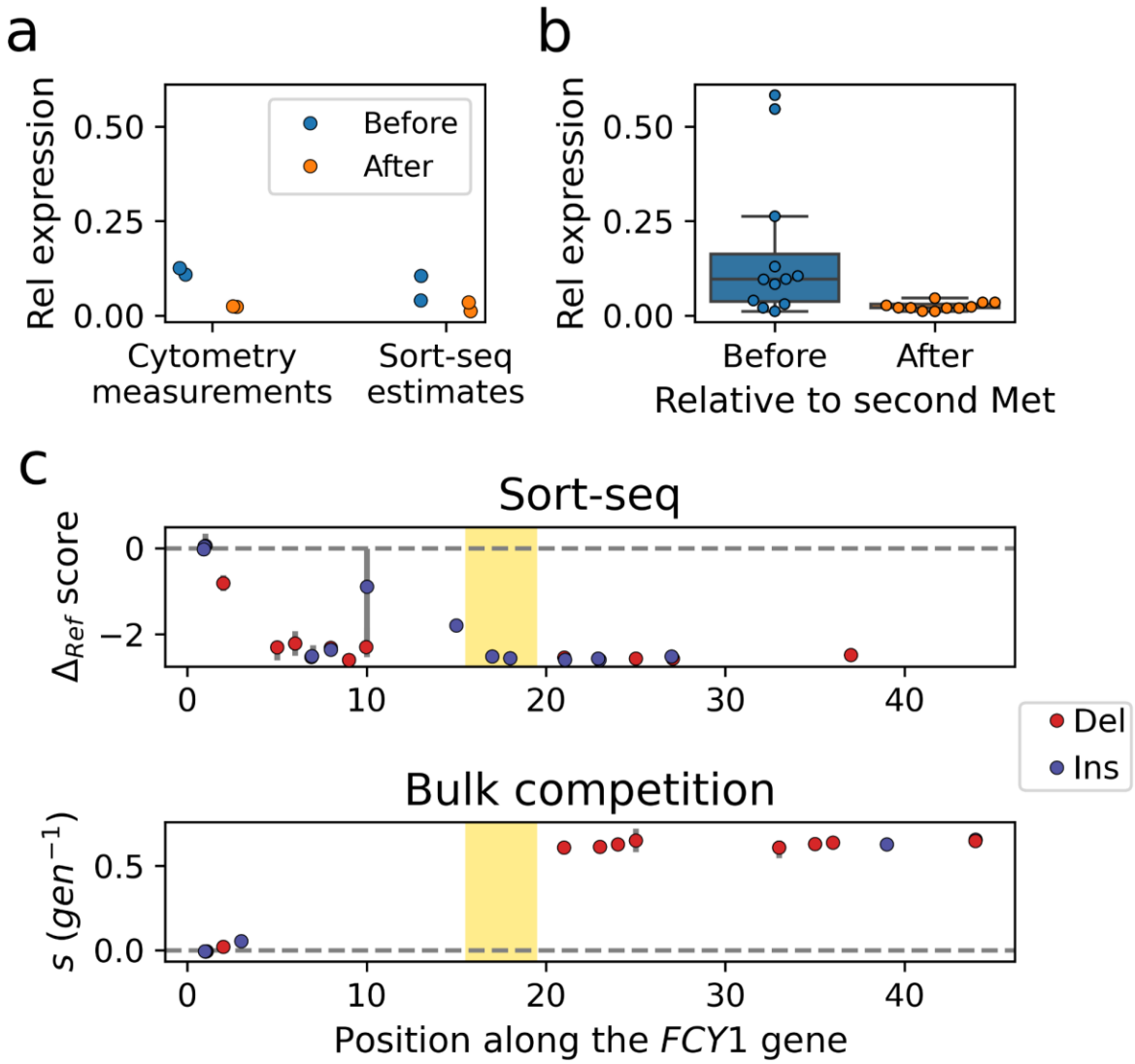
Extended Data Fig. 3: **Coding sequence mutations included as positive controls are in agreement with previously reported variant effects.** All mutants which were added to the "+*FCY1*" samples were classified into resistant (R), and sensitive (S) variants according to their effect in 5-FC conditions, using Gaussian mixture models (*Methods*). This was done using either the FC score reported by [6] 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 [6], none appear sensitive to 5-FC in the current study. Mutated sequences were collapsed into amino acid substitutions (as in [6]), resulting in 311 variants from 960 nucleotide mutations.



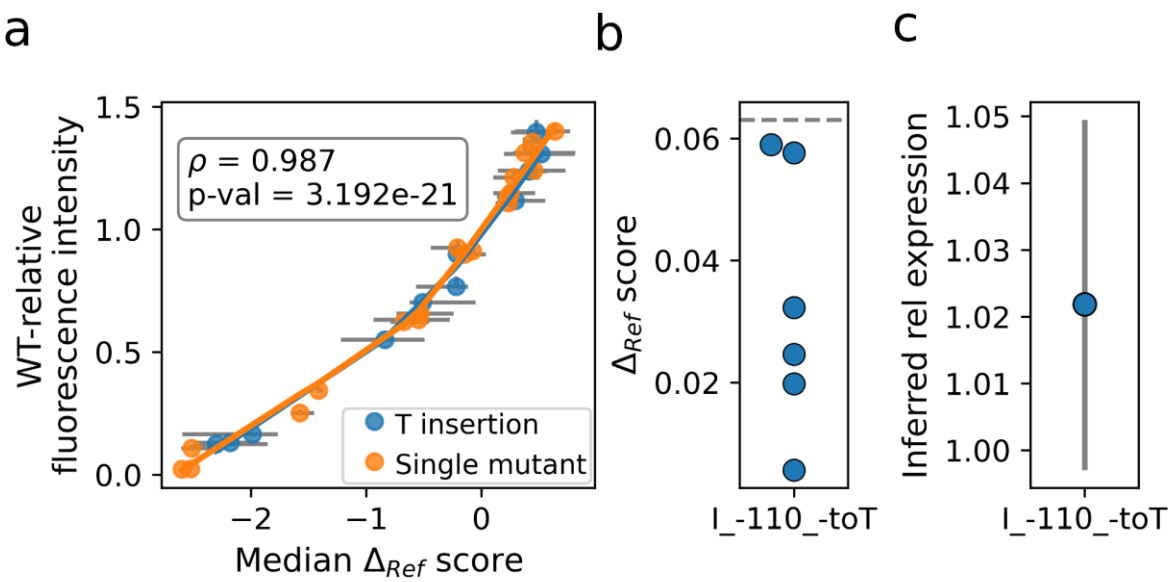
Extended Data Fig. 4: **Mixed cultures of promoter and coding sequence variants were quickly taken over by the *FCY1* mutants. a,** Gel electrophoresis image of the first PCR amplification step to prepare sequencing libraries of the F3-F4 promoter region (*Methods*), for the eight *+FCY1* cultures. Primers were designed to be specific to the pool of promoter mutants, to avoid amplifying the WT promoter sequence of the added *FCY1* variants. On the left, T0 and T1 PCR reactions have been loaded in alternance (from sample 1A to sample 4B), while the same was done on the right for T2 and T3 PCR reactions. Almost no band is visible for the T2 and T3 samples, although the same volume of PCR reaction was loaded in all cases. This suggests a much lower quantity of template DNA, and thus a smaller population size for promoter mutants. **b,** Comparisons of inferred selection coefficients in [12 µM] 5-FC for variants in the pool of promoter mutations, across samples with and without added coding sequence mutants. Variants which are identified as "CDS" are secondary mutations in the coding sequence which occurred during construction of the F3-F4 library (see **Supplementary Fig. 1**). Median selection coefficients display high agreement between the two types of samples (left). Using the minimal *s* observed among replicate cultures (right) however reveals outlier negative values for samples with added *FCY1* mutants. This is consistent with the stochastic disappearance of some genotypes due to the decreased population size of F3-F4 promoter mutants.



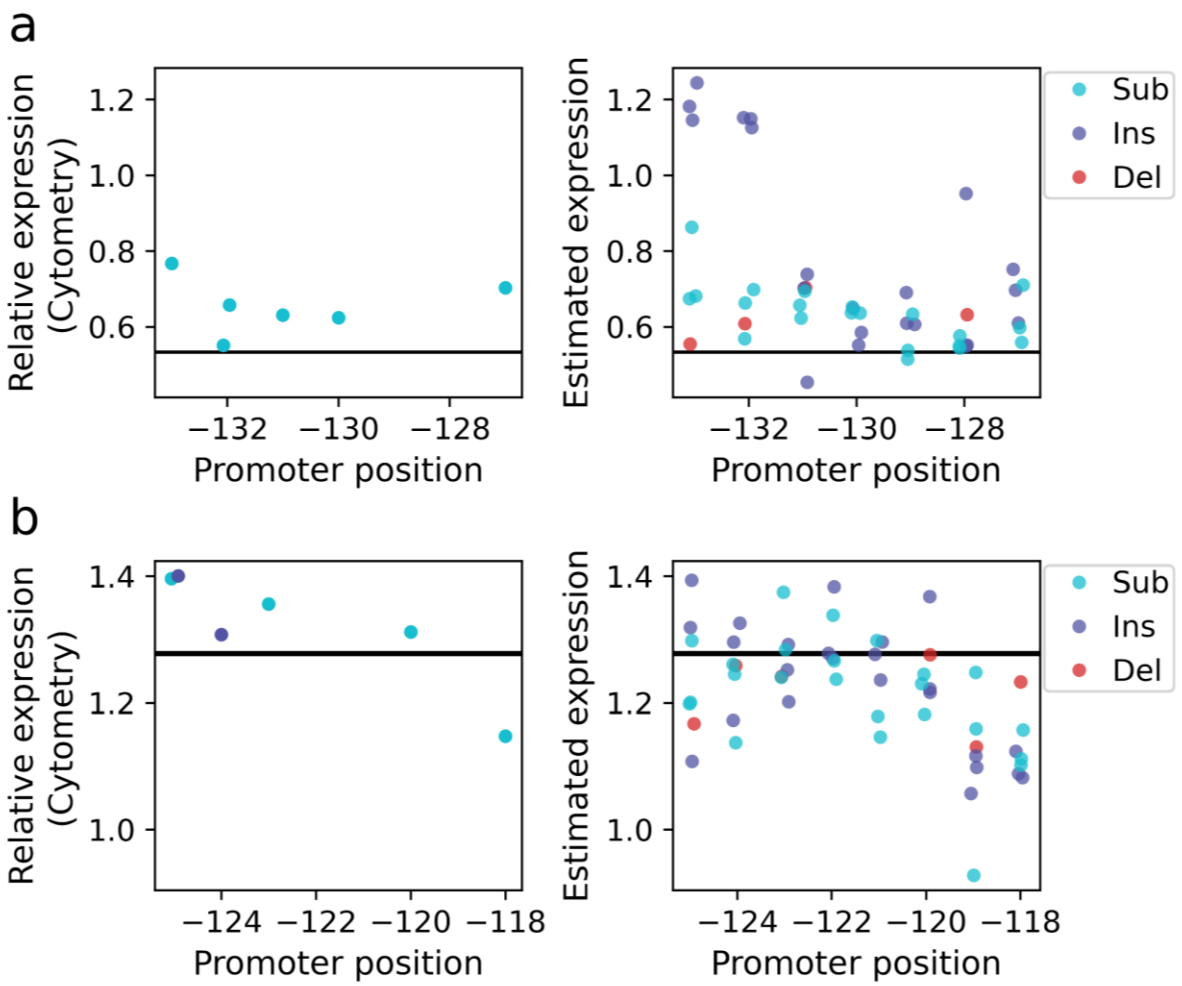
Extended Data Fig. 5: **Validation of the 42bp-mEGFP construction. a,** Medians of normalized fluorescence intensity, as measured by flow cytometry (*Methods*), for three mEGFP insertions at the genomic *FCY1* locus. Each point is a distinct clone of the corresponding strain, except for the WT. In the 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).



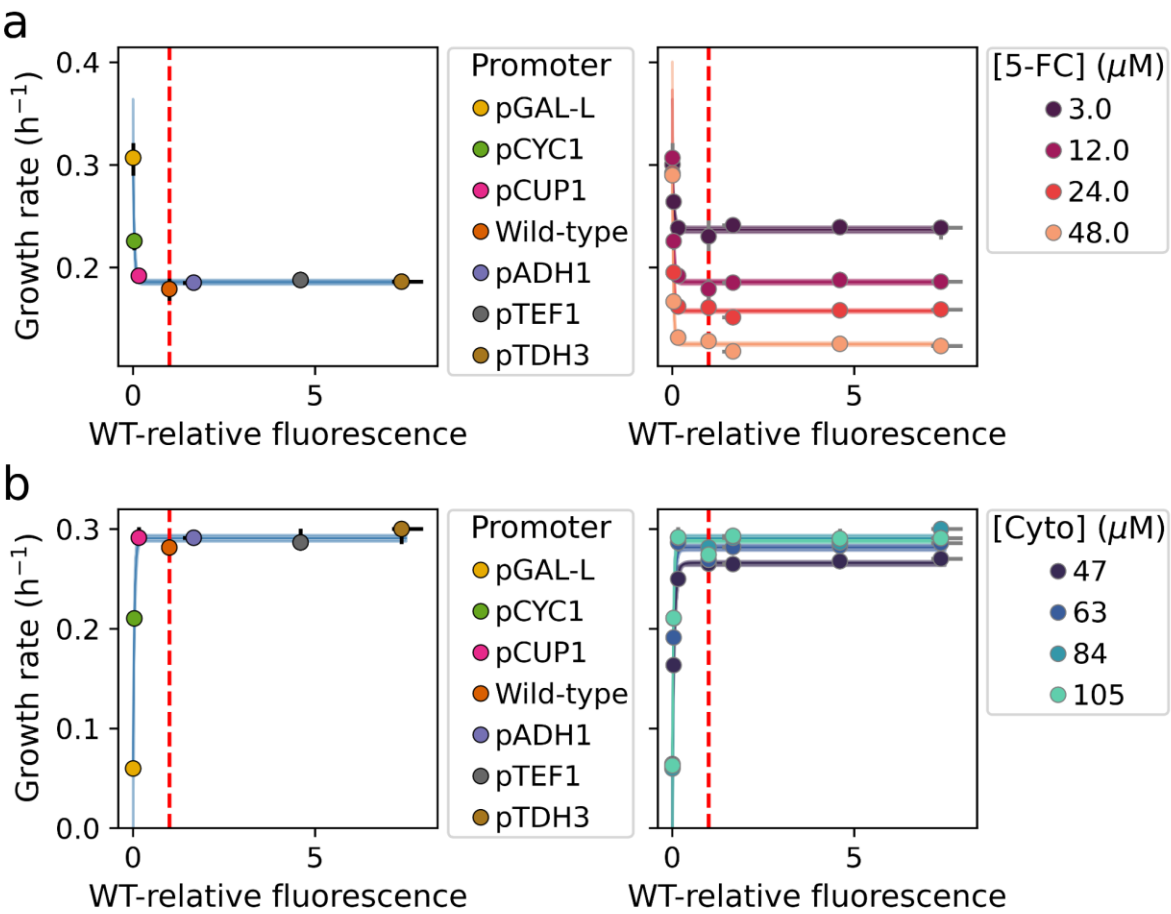
Extended Data Fig. 6: **Indels before and after the second methionine codon of *FCY1* have contrasting effects on expression level. a,** Comparison of relative expression levels for the four indel mutations which were included among the set of validation mutations, classified according to whether they 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. 7: More detailed look at the calibration curve of sort-seq scores. a, LOWESS fit as in **Fig. 4c**, but performed separately for true single mutants and for the minority of validation mutants (11/30) which featured an additional T insertion within a poly-T region (positions -110 to -99) in the F3 promoter fragment. The two fits result in almost the same curve, thus showing that these unwanted double mutants do not have a systematic effect on the conversion of Δ_{Ref} scores into relative expression levels. **b,** Sort-seq scores observed across all six replicates for the corresponding insertion (I) of a single T at position -110. All values are below the median Δ_{Ref} observed across the six sort-seq experiments (dashed line). **c,** Relative expression level inferred from the sort-seq scores shown in **b** (**Fig. 4c**), with error bars displaying the corresponding 95% prediction interval. Panels **b** and **c** show that the insertion alone has a weak to nonexistent impact on *FCY1* expression, thus confirming the robustness of relative expression estimates to the (accidental) inclusion of double mutants among the validation set.



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 [84 μM] (left), a similar plateau is observed from ~0.18 WT-relative expression. Any further increase in *FCY1* expression does not increase growth rate, such expression level variations around the WT have no fitness effect. 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).