SUPPLEMENTARY MATERIALS AND METHODS

I. Determining the false discovery rate for the detection of gene expression change

We found that independent transformants of the same promoter construct often differed in their YFP to CFP ratios due to the variation introduced by the transformation process (Gertz et al. 2009). This transformation variation does not follow a Gaussian distribution, so we chose the median YFP/CFP ratios of the 6 clones to represent the expression level of the mutant promoter. To determine which expression changes in our mutant library were statistically significant, we analyzed 186 independent transformants from a wild -type *GAL1* promoter construct to select our threshold for statistical significance. We chose 6 successive strains as one set (in total (186 – 6 + 1 = 181 sets) and used the median value of expression ratio to indicate the expression level of the group. For the reduction in gene expression, we used 10% decrease as a threshold. Since there was no instance that the median value of 181 sets was less than 10%, we claimed that using the median level of 6 biological replicates for each transformation event, the threshold provided statistically significance at a probability level of less than 1/181 = 0.5%. From this control experiment, we also found that most of the transformation variation came from the increase of expression level, probably due to the nature of the transformation process, such as multiple insertions. Even when we used a 20% increase as the threshold, the probability could only reach at the 21/181 (12%). In the mutation library, any mutation, whose median expression level of 6 biological replicates was greater than 20%, was retransformed and 12 replicates were analyzed. From this validation, we did not identify any mutation of which the median level was greater than 20% (Box 2005) (Supplementary Fig. S3).

II. Estimating the sensitivity of the expression detection system

To determine the sensitivity of our method, we examined the rate at which we detected known functional sites in the *GAL1* promoter. There are 31 bases contained in the four Gal4p binding sites and the TATA box that serve as a “test set” that we can use to quantify our false negative rate. In our study, two of these bases were not mutated or were mutated back to the consensus, so there are 29 mutations for which we expect to observe a change in gene expression. Of these, 26 displayed a significant change in gene expression, including several subtle

expression changes (~15%). From this analysis, we estimate that the sensitivity of the method is 90%.

III. Standard deviation of gene expression measurement

In addition to using median value from 6 transformants to represent the expression level, we also analyzed the standard deviation of the expression. To reduce the effect caused by transformation variation, the standard deviation is calculated by the expression value from 3 transformants, whose expression values are closest to the transformant with the median value (Supplementary Fig. S4). This analysis demonstrates that for nearly all of the mutants, the expression levels are consistent in at least 3 transformants.

In our original assay, a small subset of transformants (~10 mutations) displayed an unusually large standard deviation, so we repeated the transformation, and, for each mutant, selected an additional 12 clones to re-evaluate expression.

IV. Comparison between mutation library and natural variation

By comparing *GAL1* promoter sequences from 37 *S. cerevisiae* strains (Liti et al), we identified natural variations at 41 positions. Our library includes mutations at 37 of these positions, and mutations at 8 of them reduced gene expression by more than 10%.

V. The effect of different reading frames of uAUG on gene expression

We compared the mRNA and protein levels between genes containing an in-frame uAUG and genes containing a frameshift uAUG. In general, genes with one frameshift uAUG displayed a larger decrease in expression than genes with in-frame uAUGs (an 18% decrease vs. 10% in mRNA level, and 2.3 fold decrease vs. 2.0 fold in protein level).

VI. The features of *GAL1-10* regulatory region

We estimate that 35% of the nucleotides in the *GAL1* regulatory region are evolutionarily constrained, which is lower than the constraint observed across all yeast intergenic regions (43%) (Doniger and Fay 2007). Our system has allowed us to identify the nucleotide bases in the *GAL1* regulatory region that cause small gene expression changes when mutated, including nucleotides involved in the fine-tuning of gene expression. We found that only 7% of the bases in the *GAL1* regulatory region cause an expression change upon mutation. Several factors could explain the discrepancy between the number of observed functional bases and the number of evolutionarily constrained bases. First, our system only detected expression under one condition—galactose induction; whereas regulatory sequences respond to multiple environmental stimuli with a range of time scales. For example, mutations at the MIG1 TFBS of the *GAL1* regulatory region, which primarily responds to glucose conditions, showed no expression variation under our experimental condition. Finding the right environment under which a conserved element should be examined is challenging, as illustrated by the knock out of four "ultraconserved" elements in the mouse genome, which led to no significant phenotypic changes under any of the tested conditions (Ahituv et al. 2007). Second, our mutations were limited to one type of substitution at each nucleotide. Data inferred from PWM analysis suggests that different types of substitutions can generate a range of expression variations. Finally, *GAL1* and *GAL10* genes are divergently transcribed from a common intergenic region. Since we only examined expression for the *GAL1* gene, nucleotides that contribute to the regulation of *GAL10* (i.e. the TATA box at positions -556 to -578) would not be expected to cause a change in gene expression in our system.

VII. Intergenic sequences and primer sequences for introducing uAUG point mutations into the 21 genes

- Both wild type and mutant constructs share with the same forward primers.
- The yellow highlighted bases represent the region flanked by primers.
- The purple bases indicate single point mutations that create uAUGs.

1. CHA1 (YCL064C) chrIII:16881-17289 size: 409

TAATCGATGTGTCCTTGTTTCCTTATGCGAGCTGTTTCTTATCTATCTTATGGTCCCATTCTTTACTGCACTGTTTACATTTTGATCAATTGCGAAATGT
TCCTACTATTTTTCTTTTTCTCTTTTCGCGAGTACTAATCACCGCGAACGGAAACTAATGAGTCCTCTGCGCGGAGACATGATTCCGCATGGGCGGCTCC
TGTTAAGCCCCAGCGGAAATGTAATTCCACTGAGTGTCATTAAATAGTGCCAAAGCTTTATCAAATTGTTTGCGATGAGATAAGATAAAAGGGACAATAT
GAGGAGGAACACAGGTATATAAATATCGCCAAATAAAAGGAAAATGTTTATACAGTTTTCTCTTTTTTAAGTGCTGGATAGACAAGAGACAGGAAAATTA
ACCAGCGAG

YCL064C_Forward	TAATCGATGTGTCCTTGTTTCCTTATGCGAGC
YCL064C_Wild_Reverse	CTCGCTGGTTAATTTTCCTGTCTCTTGTCTATCC
YCL064C_Mutant30_Reverse	CTCGCTGGTTAATTTTCCTGTCTCTTGTCCATCC

2. ARG4 (YHR018C) chrVIII:141396-141886 size: 491

TCTTTAGTTACTGAAGAGTACGTGAGCGCTCACATATATACAAATATTTATACCGATTAATATTTACGTTCCTCCCTCTCTCTAATTATTCATTGATTTA
TTCAAGAATTAGCGTTATAACAATAAATGGTTGGCGCAGGCAATTAATTTTTCTTTACTCTTCCAAACCCTCTGTTAACGACAATCAAATAACCTGATCT
GCCAAGGCTCCATCATATCTGGCCTAGAACAGTTTTTTTTTTTCGATTATTTGTTCGTTCTTGTGGTGGTTACTCATTGGCAGAATCCCGAAAATCATGA
TTAGTAGATGAATGACTCACTTTTTGGATAAGCTGGCGCAAATTGAAACATGTGAAAAAAAAAAAAAAAGGATTATAAAAGGTCAGCGAAGCACAGAACTC
TGAGATAAGACTACCTTTCTTTAGCTAGGGGAGAATATTCGCAATTGAAGAGCTCAAAAGCAGGTAACTATATAACAAGACTAAGGCAAAC

YHR018C_Forward	TCTTTAGTTACTGAAGAGTACGTGAGCGC
YHR018C_Wild_Reverse	GTTTGCCTTAGTCTTGTTATATAGTTACC
YHR018C_Mutant20_Reverse	GTTTGCCTTAGTCTTGTTACATAGTTACC

3. GCD11 (YER025W) chrV:204963-205249 size: 287

GAACGCTCTTTGTTTATCTATTTATTACTAGCATTATGCGTAAGCTTGGCGTGATGTGATATATACTTGTTATATGAAGTATATCAATATGCTTTAATTT
ATGCTTATTATTTGCTCTCATGTCAGTAACCCGGACCAGATTTGCCATTGAAAAAATTTTCATAGAATAGCCCGATGCTGATCATCAGTACAAAAGACAG
AGATTTCATCTCGACCTGTACCAGGAGCTAAGCAAAGAGGGCGGAAAGAGCGTAATACACCGTTAACATCGCGCATTAGAGGTAGAC

YER025W_Forward	GAACGCTCTTTGTTTATCTATTTATTACTAGC
YER025W_Wild_Reverse	GTCTACCTCTAATGCGCGATGTTAACGGTG
YER025W_Mutant18_Reverse	GTCTACCTCTAATGCGCCATGTTAACGGTG

4. UTP10 (YJL109C) chrX:217306-217700 size: 395

GCTCTCACAAGGTCCACTTCGTTAGAAACTTGTTAACCAATGGACCAAGTGTTGCCATTGGGTCATTCTTCAGTATCAAAGAAGACTTTGACCCACTAAG
AAAAATTTCAAAAGTAAATTATAAATATAGCATTAAGTACAAATAACATATCAGCTTATGAATCCTTTTTTTCAATTTTATGGTTTCATAGCCAATAATC
ATAACATACATAGATTAATAATATTCAAGTCACATGACTTTTTTTTTTTTCTTTTTTCTTTCTGCGTTAAGCAAAGAAAAAAATTTTCAGTTAGCGATGA
GCTAGATATTTTCGAAGGTGTAACAAATACTTTCTTTATTACCTTTCAGAATCAGACTTATCTAACTGGCCACCAAAGTACGAATAACTTTGAAA

YJL109C_Forward	GCTCTCACAAGGTCCACTTCGTTAGAAAC
YJL109C_Wild_Reverse	TTTCAAAGTTATTCGTACTTTGGTGGCC
YJL109C_Mutant19_Reverse	TTTCAAAGTTATTCGTACATTGGTGGCC

5. HCA4 (YJL033W) chrX: 383371-383831 size: 461

ATAAAATAGTTAAAAATTTTTGCTGCTGGAAGCTTCAAGGTTGTTAATTTATTGACTTGCATAGAATATCTACATTTCTTCTAAAAATACATGCATAGCT
AATTCAAACTTCGAGCTTCATACAATTTTCGAGGAGATTATACTGAGTATATACGTAAATATATGCATTATATGTTATAAAATTAGAAAGATATAGAAAT
TTCATTGAAGAGTATAGAGACTGGGGTTAAGGTACTCAGTAACAGTGTCATCAATATGCTAATTTTGCGTATTACTTAGCTCTATTGCGCAAATGCAATT
TTTTCTTACCCTGATAATGCTTTATTTCCCGTTCCGAAAATTTTTCACTGAAAAAAAAGTGCTTAAGCTCATCTCATCTCATCTCATCCCATCACTATTG
AAATATTTTGCTAAAACATTATAACAGAGAGAGTTGAAAGGCTCGAGAACCTAATACTGAA

YJL033W_Forward	ATAAAATAGTTAAAAATTTTTGCTGCTGG
YJL033W_Wild_Reverse	TTCAGTATTAGGTTCTCGAGCCTTTCAACTCTCTC
YJL033W_Mutant28_Reverse	TTCAGTATTAGGTTCTCGAGCCATTCATCTCTCTC

6. PRP43 (YGL120C) chrVII: 283944-284447 size: 504

ATCGAACGATATCTTAGCCGAACAGGTCCCAATAAGGATGCCAAGGAAGCGTACTGAGTACAGTACCAGAAAGATCTTAACTAATTAAGATCATCTAAAA
AAGTACTGGGTACTATTACAGATCTAAAGCTTTACAAATGAATATTACCCGGATCATAGTTTGCTGACTGATAGTATATAGTTATGGACAAATTTCATCA
TTTTCCAAACTAAGGTCGTTTAGTTTTACGCTCTGTATTCGTTTGCTGTTTCAACTGACAGTATCATTTCGGAGTCCCATAAATCTACCAGATTGCTTTT
TATTGATCTAGCGATCTTATACACAAGTCCTTTTTTATTTCTTTTTCTTTTTTTTCGGTGCCTCTGAAAAATTTTCAAATGCGAGATGAGGTGAAAAGAA
ACTTATTCAATTGTAATATTTAACACGTTTACGGGGATGCGATTGAGTAACGATAGTATAACCTTATAAACGCGCATAGAAAGATTAGGACTGCAAGAAT
AATA

YGL120C_Forward ATCGAACGATATCTTAGCCGAACAGGTCC
YGL120C_Wild_Reverse TATTATTCTTGCAGTCCTAATCTTTCTATGCGC
YGL120C_Mutant27_Reverse TATTATTCTTGCAGTCCTAATCTTTCCATGCGC

7. RER2 (YBR002C) chrII:242571-242810 size: 240

CGTCCCACTGTACTCAATTCGGACTGAATGGCACACTTCTTGTGTTGTAGAAAATCTCTATAGATTATAATGTGTCTTACTTTAATCTTCTAATCCCCTA
ACCCTGTTACCCTCGCTCGGGCTTTTTCCGAAGATGCCAACAAAATGCTTAGAATGAAAAAAAAAAAAATTTTCGTAATAATGACAGGTTTCAATAGAAC
TTAAGCAGTAAAAATAGGGTAAACACAGGTAAAAGACGGT

YBR002C_Forward CGTCCCACTGTACTCAATTCGGACTG
YBR002C_Wild_Reverse ACCGTCTTTTACCTGTGTTTACCCTATTTTTACTGCTTAAG
YBR002C_Mutant37_Reverse ACCGTCTTTTACCTGTGTTTACCCTATTTTTACTGCATAAG

8. GCD2 (YGR083C) chrVII:646820-647126 size: 307

TACCAAAACACCCCATAGGAAATATATAAATAAACATACCTGTAAATACTAAAAAGTAGAACACAGTTCGTCACATATATTGCTATTATTGTATTAATTA
TGTTTATATTTACCCGGGAATCTCTTCACAATTTTCAGTACCTCTCGCTTTTTTCCAATTTTTCAATTTTTTTTTTTTCGAGTTTAGTGAAAATGAGTATT
CTACAACAAGGCATCACCAGCTATAGATAATCAGCGTAAGTGCATAAGATCTTATTTTCACCACGTAAGAAATACGTTGTCAAAGAAGCGTCGATAATCA
AATTGCT

YGR083C_Forward TACCAAAACACCCCATAGGAAATATATAAATAAAC
YGR083C_Wild_Reverse AGCAATTTGATTATCGACGCTTCTTTGACAACG
YGR083C_Mutant24_Reverse AGCAATTTGATTATCGACGCTTCATTGACAACG

9. ILV1 (YER086W) chrV:328137-328472 size: 336

TATGTCTGATGAATATGAAAACCTTTTCCTGACTACCAAGACTCTTTAACTCTTCTCTCTTTATTGCATATTATCTCTGCTATTTTGTGACGTTCAATTT
TAATTGACGCGAAAAAGAAAAAATAAGAAGGGCAAAAAGAAAAAGCGCAGCGGGTAGCAAATTTGGAATCGCATAAAAAGAAAAAAAAAATATCAAAGAA
AAAGAGTCATCTCAAACATATGTCTGCAGATACTTCATTATCAGCTTTGAAAACTTTTGTTGTTGCTGCTTTGAGTTCTTTCTTGTGTGAGTGCTACAAG
CCACATTTAAACTAAGTCAATTACACAAAGTTAGTG

YER086W_Forward TATGTCTGATGAATATGAAAACCTTTTCC
YER086W_Wild_Reverse CACTAACTTTGTGTAATTGACTTAGTTTAAATG
YER086W_Mutant22_Reverse CACTAACTTTGTGTAATTGACATAGTTTAAATG

10. GET3 (YDL100C) chrIV:283177-283418 size: 242

TGTAACTATATGATAAACTTGCTCAATTCAGGTTGGTTCAATGCAAATTATTACATCCCTTTTTTATCACAAGATCTTCGTAACAATAGTACTAATTAAT
AGTGCTGGGGATTCTCAACCATTTTTTTTTTTTCAAAAGCGAAAATGGATAAAACGGTGGCAACTTCAAACAAGTTGAGGGAAGCCAACGCAGCGTTTAG
GAAAACGTACGACAAGAACAAGAAGATCATCACATTGTAATT

YDL100C_Forward TGTAACTATATGATAAACTTGCTCAATTCAGG
YDL100C_Wild_Reverse AATTACAATGTGATGATCTTCTTGTTCTTGTCG
YDL100C_Mutant22_Reverse AATTACAATGTGATGATCATCTTGTTCTTGTCG

11. CDC10 (YCR002C) chrIII:118347-118617 size: 271

AGTTTGTTTGCTGATATTCCCTGGTGTTTTTGTCTCCAAGTTAAGGTATTATTTGGGAACATTGTGCCTTATATTAAGATTTGCTTTCTGAATGCTTGCG
GAAATAATCGTATGATACGATGAGATTTCCTCGAGAGACCACAAATAGATCAGAATCGCTGAATAGAATATATGTAAACATGTATCAGTAATACTTAACT
TTTTTCAGGCAAAGACAAGAAAATACAAGGCCAAGCCCCACGGTTACTACAAGCACTCTATAAATATATTA

YCR002C_Forward AGTTTGTTTGCTGATATTCCCTGGTGTTTTTG
YCR002C_Wild_Reverse TAATATATTTATAGAGTGCTTGTAGTAACCGTGGGGC
YCR002C_Mutant31_Reverse TAATATATTTATAGAGTGCTTGTAGTAACCATGGGGC

12. SNA4 (YDL123W) chrIV: 241198-241417 size: 220

ATACCGGATAATTTATGTACGAGTCATATTTGTTTTGTATAACTTTATTAATCGACTTTTTTTGTCAGGTTTATCACCCCGTTACCCATCCCGCCACTTT
TCAGCATCGAGGAAGAAAAAGGGCTGTACATTATATAAAAGGCCATCAAAAAATACCGGAAGAGGGCATCTCCCCAATAAAGGTGCCACACTGAGCAGA
GACAGATAGGGACGCGCATC

YDL123W_F ATACCGGATAATTTATGTACGAGTCATATTTG
YDL123W_wR GATGCGCGTCCCTATCTGTCTCTGCTCAGTGTGGC

YDL123W_mut_29R GATGCGCGTCCCTATCTGTCTCTGCTCATTGTGGC

13. SHR3 (YDL212W) chrIV:77968-78426 size: 459

TTTAGCTGTAGACGTCTTCAATCTCTATCCTGTCCTCTACATTAGATTACTTTCTCATTTTCTCAGAAAACCTTTTAAGCTCATCGATTAAGAAGAAAAA
AATCTTAGAAAATTTTCAGGAATGCCGGGTAAACGTATGGGGAAAGGACTTGGTGAATCGAGTTAATGTGGAGTTCGTAATAATAGTACTCTCATACCAA
GTTTTCCATCCATAATTGTTTTTGGTTGCCGTTACTTTACACCATCACCTTTCGCCAATGTAATTGGTGGCTCTTTCCCTCTCCATCTTTTTTTATAATG
TACTAAAATTAATACACGATAACAATTTTTTCGGGTTGTGTAAAAACTCTGAAAAATATAAAAAGTTGATGAAAGGACAGATTTAAGTATTTTGGAAGAC
AGCCGTAAATTCATTGGACTACAGAGCATAAAATCCCGTGATACGATAAAACTCTAACG

YDL212W_Forward TTTAGCTGTAGACGTCTTCAATCTCTATCC
YDL212W_Wild_Reverse CGTTAGAGTTTTATCGTATCACGGGATTTTATGC
YDL212W_Mutant16_Reverse CGTTAGAGTTTTATCGTATCATGGGATTTTATGC

14. GCD14 (YJL125C) chrX:186678-187124 size: 447

GAAAGAAAAACACAATACTACCAACATATATCCTGGAGTAAACTCTCAAAAATTTCCACATTCAGTTTTACCTATCCGTACACTTAATCAACTGCCTTTT
GGAGATGGAACAATAAAAAAAATGCAAGAGAGGAGGCCGTATTTGTATATAAATTAATATATTTAAATGTAAAATATATACAGTACTTGAGTTTTTTTAT
GAAGTAACGTATTTGCGCAGCATTCTTTTTCCTCTTACACTTCGTTATAACATTAACACTCATGCTATTGATACTTTTTCTTTATTATTATCTATAAAAT
TTTTTCGCTTTTTGTTTATGCCATGCCGTTAGCAACGAATTTTTTTCCATGCTCATCACTTTTAAAACTGTTTAAAGTATATAGAAAATATATGTGAATA
AGATCGGGGTGGGACTATCAGGTAGCTAAGATTACGCAGATTTAGAA

YJL125C_Forward GAAAGAAAAACACAATACTACCAACATATATCC
YJL125C_Wild_Reverse TTCTAAATCTGCGTAATCTTAGCTACCTG
YJL125C_Mutant19_Reverse TTCTAAATCTGCGTAATCATAGCTACCTG

15. SPT16 (YGL207W) chrVII:98590-98972 size: 383

TCCGGGATTCCTACTTTTTAAAGGCGTATTGTATATATTTTTGGCAGCTATGGCACACTAATTTCTATTTCCAATTTAGTGTAAGCAGAGATATAGAAGC
ATATAACATAATGATTTACTAATGGGCTCAGTACTCTCAATGCTTATTGTTATCACAAACCACGATACTAATACCCGGAAGTCGCGTGATCTTTTTTTTT
TTCGGAAAATGAGAATCGCCACTTTTCTAACGAAAGAATATTGAAATCTTAATATTGGTGAGAAGAGGAATAAAAGTAGCTAAGAGAGTCACGGGACTAA
AAGGTAAAAAGTTCAGCTATTCCAACTTTACTCCTTTCTGATATTGCACCTTTCGTGAACTGGCTTCGTGTATTTAAGTGAAT

YGL207W_Forward TCCGGGATTCCTACTTTTTAAAGGCG
YGL207W_Wild_Reverse ATTCACTTAAATACACGAAGCCAGTTCACGAAAGG
YGL207W_Mutant29_Reverse ATTCACTTAAATACACGAAGCCAGTTCATGAAAGG

16. RNA15 (YGL044C) chrVII:417041-417486 size: 446

GATGAATGCTCTTATGCGGACTGACTACACTGTGAATGTTCTGCTGCTTGAGTAATTTTTCATTCAATATTTTTCAATACGTAGTCCCATCATCGAGTGA
TCACTTGGTCATATCAGCACAGTGGGGCGGGTAAGAGTTGAAAATTTCCCTATAATGAAATTGGATATTATATATCAAAAAGGCGTTGTCATTCGAAGCA
TCGTTTCAAAAGAAGGTCCGTAAAGTTAAGGTGGGCTAGTGAAGTATCTCATAATTGTTACAATCACCATAAGAGAAATCTTGCGTTTAGAATTTCCTAA
GAGAACCGAGTCACTTCTAGCACTTGCTTTAAACAGAAACACTTGTAATAATTGTTAGAGGGATATACAAGAAAGGCTATATACCAATAGCATCACATAG
ACACGTTAATTATCAATACATTGAAAGAAAAGACGAAAACGATAAA

YGL044C_Forward GATGAATGCTCTTATGCGGACTGACTACAC
YGL044C_Wild_Reverse TTTATCGTTTTCGTCTTTTCTTTCAATGTATTG
YGL044C_Mutant21_Reverse TTTATCGTTTTCGTCTTTTCATTCAATGTATTG

17. WBP1 (YEL002C) chrV:150014-150299 size: 286

AAATGGCGGAGTAAGATCTCTGGATGATGATAGGGGGCTCTCATTGTTTTATAGATACATATTAGTATACTACAATTAAAGATATCCCAATATCTACTTT
CGATATTTTCAGGGCGGGTTATTGAATAATGCATTACCCTAAGCTACGCCTGAAAAATATCAAACACGCTTCTTGAAAAGAGAAACCTTTAAAGAGGTGT
ATAGTCAACTGAATTGGTTATAAAGTTGAATACTTTAACACAATTGACGGTGCAAATTTTGAATTATCCTTTTTTGATTATTTAGG

YEL002C_Forward AAATGGCGGAGTAAGATCTCTGGATGATG
YEL002C_Wild_Reverse CCTAAATAATCAAAAAAGGATAATTCAAAATTTGCACC
YEL002C_Mutant29_Reverse CCTAAATAATCAAAAAAGGATAATTCAACATTTGCACC

18. TIP20 (YGL145W) chrVII:229692-230247 size: 556

GTTTTCACAATGACACAGAGAAACTAGGCCTTAAAGACGGTAGAATCTGTAAGGAGAGGCAAAAACCAGAAACTTTTCATAAGATTCAGTAAAGTCTTTT
TACTACTAGAGTGTACGGCTTTTATTTTATTTTGAACACAACTACAGCCTCACCACGGTATTTTTTTTTCCTTTGAAGAAGTATCAACGACTTAGGGTGC
AATAGCATTACCCATCCCCACATTATGATGCATTATTAAACTTAACTTTACTGGCATTGGCTACCGTCGCCGGGCCGAGGGTACAAAATCGAGAAGTCAT
AAAATCAGGGTTATATGTAAAATACAGTAGGGCATAGCATTATATAACGGGGTTCTTAAATAGGGCTGAAAGAAGAATATTAGGTAAAAATTCTACATTT
CTACGAGCCCTGTTTTATTGATATTTCAATATAAATACAATAGTATAACAATAGTGTAATAATAGTAGATGAATCCTAAAAATTGTAGGCAGAAGTAGAT
AAGAAATTGTATATTAAGTTATTGTTTATAAGCATAGTCACAAGTGCATAAAAACT

YGL145W_Forward GTTTTCACAATGACACAGAGAAACTAGGC
YGL145W_Wild_Reverse AGTTTTTATGCACTTGTGACTATGCTTATAAAC
YGL145W_Mutant21_Reverse AGTTTTTATGCACTTGTGACCATGCTTATAAAC

19. TSC10 (YBR265W) chrII:738365-738576 size: 212

ATCGTATTCCTCCGTATTTTGTGCTACTTGTATTTCCTGGAGTAATTTTACTATCATTGTTCCTTAACGTGAAACGTAAGGCTAAAAAACAATAAACGAC
AAAAAAAAAACAAAAAAATGAAAAATGCAGAGGAGTTTCTCTGCATTTACTGGAAGCGATTAAATAAGATTGCCTCAAGAAAAGAAGTTATCGGTATAAT
CTTTTGTAGGTA

YBR265W_Forward	ATCGTATTCCTCCGTATTTTGTGCTACTTG
YBR265W_Wild_Reverse	TACCTACAAAAGATTATACCGATAACTTCTTTTC
YBR265W_Mutant30_Reverse	TACCTACAAAAGATTATACCGATAACTTCATTTC

20. TAF6 (YGL112C) chrVII:299732-299980 size: 249

TTTACGATCGAAGCTGGGCCTATTTATGGTTAAATTTTCTCTGTATTATGTTCATTGAAGTGAGGCTTCGAGATTGGGCATCGCTTTACAAAATTTTTCA
GCTTTATTTTATTACCAAAAAAATACCCGGAAACCATCAATAACAAACACCGATTCAAGTAGCAGCAAAAGTAGTTCATGAGCGAAAACAATAGATTGGC
TTTAAAAAGGAAGCAAAGGACGCCTCAGTGTAGTTGATAGCCTGTTAAA

YGL112C_Forward	TTTACGATCGAAGCTGGGCCTATTTATGG
YGL112C_Wild_Reverse	TTTAACAGGCTATCAACTACACTGAGG
YGL112C_Mutant16_Reverse	TTTAACAGGCTATCATCTACACTGAGG

21. YGL082W chrVII:355476-355829 size: 354

TTTGTCAGCCTCCCGAAGATGAAGATATCTTAAATTGGCTTGAAACGAACGCACCACGAAGTCAGTGGTTCTTTGTATTAGTAAATAAACCAAATATAAT
TTTGTACCTTAGCAGCTCTTTCGAGAAATTCATTTTCAGTTTAGTAGGCCTGACTGACAAGCAAATTGTTCTTTTTTCTGGACGGTGCGGGATCAGGGGA
ATCACACGTTCAATAAATAAAGAAAGGATGAAACAGTCGTACAACAACCAAACATGAAGGAGGTACAGAGGAAGCAAACAAAACCAAATTTGCTAAAATC
AAGACTGGGGCAATTTTTCAGTCCGTATACAATCGTTCACAGGAATTAGCCAAT

YGL082W_Forward	TTTGTCAGCCTCCCGAAGATGAAGATATC
YGL082W_Wild_Reverse	ATTGGCTAATTCCTGTGAACGATTGTATACG
YGL082W_Mutant21_Reverse	ATTGGCTAATTCCTGTGAACCATTGTATACG

Table S1. The spectrum of mutation library

We listed 12 substitution types in the *GAL1* mutation library, including 582 mutated nucleotide positions.

Reference sequence nucleotide type	Mutated sequence nucleotide type	Count	Frequency
A	C	54	0.09
A	G	63	0.11
A	T	80	0.14
C	A	55	0.09
C	G	30	0.05
C	T	41	0.07
G	A	50	0.09
G	C	4	0.01
G	T	43	0.07
T	A	74	0.13
T	C	37	0.06
T	G	51	0.09

Table S2. Mutation spectrum and sequence comparison among four yeast species

	Total	Conserved	Mutation found in at least 1 yeast species	Mutations not found in the 4 yeast species
Mutations that changed expression	43	35 (81%)	2 (5%)	41(95%)
Mutations that did not change expression	538	210 (39%)	90 (17%)	448 (83%)

SUPPLEMENTARY REFERENCES

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, and Rubin EM. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5:** e234.

Box GEP. 2005. *Statistics for Experimenters: design, innovation and discovery*. John Wiley & Sons, Inc. , Hoboken, New Jersey.

Doniger SW and Fay JC. 2007. Frequent gain and loss of functional transcription factor binding sites. *PLoS Comput Biol* **3:** e99.

Gertz J, Siggia ED, and Cohen BA. 2009. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457:** 215-218.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V et al. 2009. Population genomics of domestic and wild yeasts. *Nature* **458:** 337-341.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Distribution of point mutations per clone for the *GAL1* mutation construct library created by linear random PCR mutagenesis method. This is a histogram of clones with different point mutation occurrences. The value above each bar is the number of counts in that category.

Figure S2. The distribution of nucleotide positions of the *GAL1* promoter based on the classification of mutation type. At any given nucleotide, one out of five types of mutation could occur; indicated by Mut. Type 0 to 4: Type 0 = No mutation; Type 1 = 1 out of 3 types of substitutions; Type 2 = 2 out of 3 types of substitutions; Type 3 = 3 out of 3 types of substitutions; Type 4 = 3 types of substitution and insertion/deletion type.

Figure S3. The distribution of the expression ratio for control strains to determine the sensitivity of the expression detection system (see Supplement methods). This histogram shows the distribution of the median expression ratio of 6 independent yeast transformants from 181 sets of biological replicates. The vertical purple line, at ratio 0.9, has been used as the threshold (>= 10%) to determine the down-regulation of the *GAL1* gene, with p value <0.005.

Figure S4. Standard deviation of gene expression for the 582 mutated bases at position -1 to -630 in the *GAL1* promoter. The X-axis labels are the wild-type *GAL1* promoter sequence.

Figure S5. Mutations at the 5'UTR region and expression changes upon mutation. The abbreviations and the color bar indicate expression change (see fig. 2 in the main text). Three red circles represent positions 16, 36, and 45, respectively. The pink boxes represent the creation of uAUG by a single point mutation at these positions.

FigureS6. Ratio of the observed and expected incidences of 64 trinucleotides computed by both zero order and first order Markov chain models, showing uAUG as the most depleted trinucleotide in both models.

Figure S7. P-value ranking of the hypergeometric test for 64 trinucleotides. For each trinucleotide, the statistical test examined the significance of overlapped gene sets that contained each trinucleotide between *S.cerevisiae* and *S. paradoxus.* The plot shows that the ATG trinucleotide has the most significant p-value score.

Figure S8: (a) The distribution of the distance for the first uAUG sites relative to the canonical ATG start site in *S. cerevisiae* genome. (b) The distribution of the length of the 5' UTR in *S. cerevisiae* genome.

SUPPLEMENTARY TABLE LEGENDS

Table S3. Gene Expression Variation for *GAL1-10* Regulatory Region. Position from -1 to -630 bp.

Table S4. GenBank Accession Numbers.