

Supplementary Information for
Massively Parallel Single Amino Acid Mutagenesis

Jacob O. Kitzman, Lea M. Starita, Russell S. Lo, Stanley Fields, and Jay Shendure

1 **Online Methods**

2 **Mutagenic primer preparation**

3 Mutagenic primers were electrochemically synthesized on a 12,432-feature
4 programmable DNA microarray and released into solution by CustomArray, Inc²⁵. For Gal4 (GI
5 #6325008), codons 2-65 were each replaced with the optimal codon in *S. cerevisiae*
6 corresponding to one of the 19 other amino acids²⁶, a stop codon (TAA), or an in-frame deletion,
7 for a total of 1,344 oligos, each synthesized in duplicate (for a total of $2 \times 64 \times (19 + 1 + 1) = 2,688$
8 array features). For p53 (GI #120407068), codons 1-393 were replaced with fully degenerate
9 bases (“NNN”) during synthesis, such that primer molecules synthesized within a single spot on
10 the array are degenerate for the triplet corresponding to a single residue, for a total of 393 oligos,
11 each synthesized in triplicate (for a total of $3 \times 393 = 1,179$ array features).

12 Each primer was designed as a 90mer, including flanking 15-base flanking adaptor
13 sequences, except for the Gal4 in-frame codon deletion primers, which were designed as 87mers.
14 Each primer is synthesized sense to the gene, with 33 upstream bases, followed by the codon
15 replacement, and 24 downstream bases. To allow for specific retrieval, a different flanking
16 adaptor pair was used for each subset of mutagenic primers on the array. Gal4 primers were
17 flanked by adaptor sequences “truncL_GAL4DBD” and “truncR_GAL4DBD” and p53 primers
18 were flanked by “truncL_TP53” and “truncR_TP53” (**Supplementary Table 6**). Mutagenic
19 primer libraries were retrieved by PCR using the respective adaptor pair (“L_TP53”/“R_TP53”
20 or “L_GAL4DBD”/“R_GAL4DBD”), using 10 ng of the starting oligo pool as template using
21 Kapa Hifi Hot Start ReadyMix (“KHF HS RM”, Kapa Biosystems) and following the cycling
22 program “ADO_KHF” (**Supplementary Table 7**). Reactions were monitored by fluorescent
23 signal on a BioRad Mini Opticon real-time thermocycler, and were removed after 15 cycles.

24 Amplification products were purified with Zymo Clean & Concentrate 5 columns (Zymo
25 Research). Electrophoresis on a 6% TAE polyacrylamide gel confirmed a single band of ~108 bp
26 for each library, corresponding to the original oligo size plus 18 bp of additional adaptor
27 sequence added by PCR (**Supplementary Fig. 11**).

28 The resulting oligo pools were further amplified with adaptors modified to contain a
29 deoxyuracil base at the 3' terminus. This second-round amplification was carried out in 50ul
30 reactions, using 1ul of the previous amplification reaction (at a 1:4 dilution in dH₂O) as template,
31 following cycling program "ADO_KR". Each reaction included 25ul Kapa Robust Hot Start
32 ReadyMix (which is not inhibited by uracil-containing templates), amplification primers at 500
33 nM each ("L_"GAL4DBD"/"R_GAL4DBD_U" or "L_TP53"/"R_TP53_U"), and SYBR Green I
34 at 0.5X. Immediately following PCR, each library was denatured at 95°C for 30 seconds, and
35 then snap cooled on ice. To cleave the "R" adaptors, 2 U USER enzyme mix (New England
36 Biolabs) was added, and each reaction was incubated for 15 minutes at 37°C. Finally each
37 reaction was supplemented by 2.5ul of a 10 uM stock of the corresponding "L" primer
38 ("L_GAL4DBD" or "L_TP53"), followed by one final cycle of annealing/priming/extension.
39 Amplification products were purified as before on Zymo columns. Gel electrophoresis confirmed
40 that each resulting library was a mixture of off-product flanked on both sides by adaptors (108
41 bp), and the desired product with only "L" adaptors (84 bp, **Supplementary Fig. 11**).

42 **Wild-type template preparation**

43 The full-length Gal4 open reading frame was amplified from genomic DNA of *S.*
44 *cerevisiae* strain BY4741 and directionally cloned into the yeast shuttle vector p416CYC, a
45 single-copy CEN plasmid with the CYC1 promoter²⁷ by digestion with SmaI and ClaI (New
46 England Biolabs), using the InFusion cloning kit (Clontech). Subsequently, an N-terminal

47 truncation was prepared by amplifying residues 1-196 from the original clone using the primer
48 pairs GAL4_CLONE_F and GAL4_NTERM_R, and recloning into p416CYC to create
49 p416CYC-Gal4Wt-1-196. This fragment retains the same DNA-binding specificity as full-length
50 Gal4 and is sufficient for transcriptional activation¹⁴. Enforced expression of full-length Gal4
51 causes cellular toxicity by aberrantly sequestering the transcriptional machinery, in an effect
52 called squelching²⁸. A similar effect is observed for Gal4 1-196 (i.e., loss-of-function alleles are
53 more fit than wild-type ones under non-selective growth, **Supplementary Fig. 6**) but to a much
54 lesser degree than for the full length protein. For p53, a wild-type clone with a C-terminal GFP
55 fusion was purchased from OriGene (#RG200003).

56 To prepare wild-type sense and antisense strands to serve as templates for mutagenic
57 primer extension, the desired fragments were amplified from plasmid clones by PCR. To select
58 for the sense strand, the reverse primer was phosphorylated to allow for its later degradation by
59 lambda exonuclease, and to select the antisense strand, the forward primer was instead
60 phosphorylated. Furthermore, to minimize undesired carry-through of wild-type copies, in some
61 cases long synthetic tails (38 or 40 nt) were placed on the phosphorylated primer to prevent the
62 resulting 3' ends of the selected strands from acting as primers during subsequent extension
63 steps. Primers were either ordered with a 5' phosphate or were enzymatically phosphorylated in
64 10ul reactions containing 1 ul of 100uM primer stock, 7 ul H₂O, 1ul 10X T4 Ligase Buffer with
65 ATP (NEB), and 10U T4 polynucleotide kinase (NEB) and incubated for 30 minutes at 37°C,
66 followed by heat inactivation for 20 minutes at 65°C and one minute at 95°C. Wild-type
67 fragments were amplified in 50ul PCR reactions with forward and phosphorylated reverse
68 primers using Kapa HiFi U+ HotStart Ready Mix (“KHF U+ HS RM”) supplemented with
69 dUTPs to a final concentration of 200 nM. Primers for wild-type template preparation are listed

70 in **Supplementary Table 6**, and amplification used cycling conditions “WT_STRAND_PREP”.
71 For starting template, 200 pg of each wild-type clone plasmid was used. Amplification products
72 were purified by Zymo column, and to select the desired strand, 30 ng of each PCR product was
73 treated for 30 min at 37°C with 7.5 U lambda exonuclease (NEB) in a 30ul reaction containing
74 lambda exonuclease buffer at 1X final. Reactions were heat killed for 15 minutes at 75°C and
75 purified by Zymo column (5 volumes binding buffer, eluted in 10ul buffer EB).

76 **Mutagenic primer extension**

77 Next, 2 ng of each primer pool was combined with 3 ng of its respective sense-strand
78 template, raised to 12.5ul with dH₂O, and mixed with 12.5ul of KHF U+ HS RM for extension
79 along the dUTP-containing wild-type template by the annealed mutagenic primers. The reaction
80 was subjected to one round of denaturation, annealing, and extension (cycling conditions
81 “PALS_EXTEND”), purified by Zymo column, treated with 1.5U USER enzyme for 10 minutes
82 at 37°C to degrade the wild-type template, and purified again by Zymo column (same
83 conditions).

84 The resulting strand extension products were enriched via PCR using the KHF U+ HS
85 RM in 25 ul reactions using the cycling program PALS_AMPLIFY and 3 ul of preceding strand
86 extension product as template. Reactions were monitored by SYBR Green fluorescence intensity
87 and removed in mid-log phase (13 cycles for Gal4, 10 cycles for p53). The forward and reverse
88 primers corresponding to the sense strand template and the mutagenic adaptor, respectively, were
89 “OUTER_F”/“L_GAL4DBD_U” (for Gal4) or “P53_SENSE_F”/“L_TP53_U” (for p53). An
90 aliquot of each amplification product was visualized by PAGE electrophoresis, and appeared as a
91 smear over the expected size ranges (~450-650 bp for Gal4, ~300-1500 bp for p53,
92 **Supplementary Fig. 11**).

93 The reverse primer in the preceding amplification step carried a 3'-terminal dUTP,
94 allowing for adaptor excision by treatment with 1 U USER enzyme for 15 minutes at 37°C. This
95 reaction was cleaned by Zymo column and eluted in 11.8 ul buffer EB. Next, the respective
96 forward primer was added (0.75 ul at 10 uM) followed by 12.5ul of KHF HS RM to create sense-
97 strand mutagenized megaprimers with one round of cycling conditions "PALS_EXTEND". For
98 this step, the non-uracil tolerant PCR mastermix was used to limit amplification of any
99 remaining uracil-containing wild-type strand template. Alternatively, adaptor sequences could be
100 designed to allow excision with Type IIS restriction enzymes.

101 Sense-strand megaprimers were then purified by Zymo column, annealed to the wild-type
102 antisense strand, and extended to form full-length copies. Each extension reaction contained 3 ng
103 of the sense-stranded megaprimer pool, 1 ng of the wild-type dUTP-containing antisense strand,
104 and was performed with KHF U+ HS RM, followed by column cleanup, USER treatment (1.5U
105 for 10 min at 37°C), and a second column cleanup, as during the initial mutagenic strand
106 extension reaction. Finally, the full-length mutagenized copies were enriched by PCR using fully
107 external primers ("OUTER_F"/"GAL4_OUTER_R" or "OUTER_F"/"P53_ANTISENSE_R"),
108 in 25 ul PCR reactions with KHF U+ HS RM with conditions "PALS_AMPLIFY".

109 **PALS library cloning**

110 Gal4 DBD PALS libraries were cloned into p416CYC-bc, a pre-tagged library of vectors
111 derived from p416CYC, in which each clone contains a random 16mer tag. To prepare
112 p416CYC-bc, a pair of unique restriction sites was placed downstream of the CYC1 terminator
113 by digesting p416CYC with KpnI-HF (NEB) and inserting a duplex of oligos
114 ("P416CYC_AGEMFE_TOP"/"P416CYC_AGEMFE_BTM") by ligation to create the
115 following series of restriction sites: KpnI-AgeI-MfeI-KpnI. A tag cassette containing a

116 randomized 16mer (“P416CYC_BC_CAS”) was then PCR-amplified using primers
117 “P416CYC_AMP_BC_CAS_F”/“P416CYC_AMP_BC_CAS_R” and cycling program
118 “MAKE_BC_CAS”, to add priming sites for later tag counting during Gal4 functional selections,
119 and to add flanking AgeI and MfeI sites. The resulting tag cassette amplicon was directionally
120 cloned into the modified p416CYC vector by double-digestion with AgeI-HF and MfeI-HF
121 (NEB) and transformed into ElectroMax DH10B electrocompetent *E. coli* (Invitrogen), to yield
122 $\sim 9.2 \times 10^6$ distinctly tagged clones. The resulting library, p416CYC-bc, was expanded by bulk
123 outgrowth and purified by midiprep using the ChargeSwitch Pro Midi kit (Invitrogen). Next, 15
124 μ g of p416CYC-bc was digested with 40U SmaI (NEB) for 1 hr at 25°C in 60ul, followed by
125 addition of 20U ClaI (NEB), digestion for 1 hr at 37°C, and purification by MinElute column
126 (Qiagen). To insert the Gal4 DBD PALS library, 50 ng of the final PALS PCR product was
127 combined with 10ng SmaI/ClaI linearized p416CYC-bc vector and directionally cloned using the
128 InFusion HD kit (ClonTech), as directed. Libraries were transformed by electroporation into 10-
129 beta electrocompetent *E. coli* (NEB), and bulk transformation cultures were expanded overnight
130 in 25ml LB + ampicillin (50 μ g/ml) at 37°C, shaking at 250 rpm. Due to the large number of
131 vector copies present in the cloning reaction, pairing of Gal4 mutant inserts with tag is
132 essentially sampling with replacement; the number of positive clones ($\sim 9.0 \times 10^5$) is less than the
133 number of tags by approximately an order of magnitude, so only $\sim 0.45\%$ of tags are expected to
134 be paired with two different inserts.

135 Tagged p53 PALS libraries were created in the reverse order: the PALS-mutagenized
136 amplicon was cloned first, and the library was expanded and tags inserted second. The p53
137 library was cloned into pCMV6-AC-GFP (Origene) by standard directional cloning in two
138 separate cloning reactions using NotI-HF/BamHI-HF or NotI-HF/KpnI-HF (NEB). Libraries

139 were transformed into 10-beta electrocompetent cells (NEB), combined, expanded overnight and
140 purified by midiprep as for Gal4. Subsequently, the cloned p53 libraries were linearized at the
141 AgeI site downstream of the hGH poly-A signal: 2.5 ug of plasmid DNA was digested with 10U
142 AgeI (NEB) in 50ul for 1 hr at 37°C, and purified by Zymo column. A tag cassette containing a
143 randomized 20mer was synthesized (“P53_BC_CAS”) and PCR amplified for cloning (using
144 primers “P53_AMP_BC_CAS_F”/“P53_AMP_BC_CAS_R”), using KHF RM HS and cycling
145 program “MAKE_BC_CAS”. Tags were directionally inserted at the AgeI site by InFusion
146 cloning, as for Gal4, and the resulting plasmid was transformed, expanded in bulk, and purified
147 by midiprep as in the first round of cloning.

148 **Clone subassembly sequencing**

149 To bring the tag cassette into proximity with the mutagenized Gal4 coding sequence
150 (**Supplementary Fig. 10**), 1 ug of the mutant Gal4 plasmid library was digested with 20 U
151 BamHI-HF (NEB) in 1X CutSmart Buffer for 30 minutes at 37°C. The digest was cleaned up by
152 Zymo column, and 200 ng of the product was recircularized by intramolecular sticky-end
153 ligation using 1600U T4 DNA ligase (NEB) in a 200ul reaction for 2 hours at 20°C. Following
154 Zymo column cleanup, linear fragments and concatamers were depleted by treatment with 5U
155 plasmid-safe DNase (Epicentre) for 30 minutes at 37°C, and then 30 minutes at 70°C. Next, PCR
156 was used to amplify fragments containing the tag cassette at one end, and the mutagenized insert,
157 using 3ul of the heat-killed recircularization product as template (expected recircularization
158 product and primer pairs shown in **Supplementary Fig. 10a**) and following cycling conditions
159 “PALS_SUBASSEM”. Amplification products were purified using Ampure XP beads (1.5X
160 volumes bead/buffer). P53 PALS clone libraries were recircularized following a similar strategy,
161 except that digestions with EcoRI or NotI followed by recircularization were used individually to

162 bring the tag cassette into proximity with the N or C termini, respectively (**Supplementary Fig.**
163 **10b**).

164 To prepare Illumina sequencer-ready subassembly libraries, tag-linked amplicons from
165 the previous step were fragmented and adaptor-ligated using the Nextera v2 library preparation
166 kit (Illumina), with the following modifications to the manufacturer's directions: for each
167 reaction, 1.0ul Tn5 enzyme "TDE" was combined with 2.0 ul H2O, 5ul Buffer 2X TD, and 2ul
168 of the post-recircularization PCR product. Longer insert sizes were obtained by diluting enzyme
169 TDE up to 1:10 in 1X Buffer TD (a 1:4 dilution was used for the libraries sequenced here).
170 Tagmentation was carried out by incubating for 10 minutes at 55°C, followed by library
171 enrichment PCR to add Illumina flowcell sequences. Libraries were amplified by KHF RM 2X
172 mastermix in 25ul using a forward primer of NEXV2_AD1 and one of the indexed reverse
173 primers "SHARED_BC_REV_####". PCR reactions were assembled on ice using as template 2ul
174 of the transposition reaction (without purification), and cycling omitted the initial strand
175 displacement step typically used with the Nextera kit (conditions
176 "NEXTERA_SUBASM_PCR"). Lastly, fixed-position amplicon sequencing libraries starting
177 from the mutagenized insert end of the clone were prepared by adding Illumina flowcell adaptors
178 directly to the tag-insert amplicons by PCR, using the same PCR conditions but substituting the
179 forward primer "ILMN_P5_SA" for the Nextera-specific forward primer.

180 **Tag-directed clone subassembly**

181 Subassembly libraries were pooled and subjected to paired-end sequencing on Illumina
182 MiSeq and HiSeq instruments, with a long forward read directed into the clone insert (101 bp for
183 HiSeq runs, 325 or 375 bp for MiSeq runs) and a reverse read into the clone tag. Tag-flanking
184 adaptor sequences were trimmed using cutadapt (obtained from

185 <https://code.google.com/p/cutadapt/>), and read pairs without recognizable tag-flanking adaptors
186 were excluded from further analysis. Insert-end reads were aligned to the Gal4 or p53 wild-type
187 clone sequence using bwa mem²⁹ (with arguments “-z 1 -M”), and alignments were sorted and
188 grouped by their corresponding clone tag. To properly align the programmed in-frame codon
189 deletions included in the Gal4 PALS library, bwa alignments were realigned using a custom
190 implementation of Needleman-Wunsch global alignment with a reduced gap opening penalty at
191 codon start positions (match score=1, mismatch score=-1, gap open in coding frame =-2, gap
192 open elsewhere=-3, gap extend=-1). A consensus haplotype sequence was determined for each
193 tag-defined read group by incorporating variants present in the group’s aligned reads at sufficient
194 depth. Spurious mutations created by sequencing errors, or mutations present at low allele
195 frequency arising from linking two haplotypes to the same tag were flagged and discarded by
196 requiring the major allele at each position (either wild-type or mutant) to be present with a
197 frequency of $\geq 80\%$, $\geq 75\%$ and $\geq 66\%$, for read depths ≥ 20 , 10-19, or 3-10, respectively,
198 considering only bases with quality score ≥ 20 . Tag groups with fewer than three reads (Gal4
199 DBD) or 20 reads (p53) were discarded, as were groups not meeting the major allele frequency
200 threshold across the entire target (Gal4 DBD) or a minimum of 1 kbp (p53). Consensus
201 haplotypes were validated by Sanger sequencing of individual colonies from each tagged
202 plasmid library (**Supplementary Fig. 12** and **Supplementary Table 1**).

203 Subassembled clones with multiple mutations examined to investigate the underlying
204 cause of the secondary mutations. For the Gal4 DBD PALS library, these were dominated by
205 PCR chimeras (52% among clones with secondary mutations) and synthesis errors (24%), as
206 estimated by counting clones bearing two programmed mutations, or one programmed mutation
207 and secondary mutations within the boundaries of the corresponding mutagenic primer.

208 Chimerism is a technical challenge commonly encountered while amplifying libraries of
209 homologous sequences³⁰, when incomplete strand extension products in one cycle of
210 amplification act as primers in the subsequent cycle. Future optimization efforts will be directed
211 at quantifying and mitigating this phenomenon by manipulating input template concentration and
212 minimizing amplification cycles, or alternatively using droplet PCR³¹. To reduce the impact of
213 synthesis errors, PALS uses short oligonucleotides (90 nt), but it will nevertheless benefit from
214 ongoing developments in high-fidelity synthesis³². In addition, as single-base deletions are the
215 dominant synthesis error mode³³, stringently size-selecting primer libraries may further enrich
216 for primers lacking undesirable secondary mutations. Another strategy would fuse libraries in-
217 frame to a selectable marker in the bacterial cloning host, although our preliminary observations
218 suggest that such selection is inefficient for proteins that do not fold or express well in *E. coli*.
219 For p53, because codon substitutions were encoded as “NNN”, the origin of secondary mutations
220 could not be distinguished between synthesis errors, PCR errors, or chimerism between
221 fragments each bearing a single codon swaps.

222 Although PALS is intended to create single-mutant clones, for applications such as
223 protein engineering, it may be useful to obtain multiple mutations per copy. This could be
224 accommodated by applying PALS serially, to first create a library of single-mutant copies which
225 would then be used as the starting template for the second round. In the context of typical-length
226 genes, the multi-mutation space is so large (e.g., for TP53 double mutants,
227 $\text{choose}(393,2) * 19 * 19 = 2.78 \times 10^7$ possibilities) that it may be technically impractical to construct,
228 much less survey, the entire space. By serially applying PALS, however, it could be possible to
229 focus on a defined subspace using a subset of mutagenic primers in either or both rounds.

230 **Gal4 functional selections**

231 Gal4 DBD PALS libraries were transformed into chemically competent *S. cerevisiae*
232 strain PJ69-4alpha³⁴ prepared using a modified LiAc-PEG protocol, as previously described^{35,36}.
233 After transformation, cells were allowed to recover for 80 minutes at 30°C shaking at 250 rpm.
234 To select for transformants, cultures were spun down at 2000 x *g* for 3 min, resuspended and
235 grown overnight at 30°C in 40 ml SC media lacking uracil. Plating 0.25% of the recovery culture
236 prior to outgrowth indicated a library of ~ 2x10⁵ transformants. Following overnight outgrowth,
237 glycerol stocks were prepared from the transformation culture and stored at -80°C.

238 Frozen stocks of yeast carrying the Gal4 DBD PALS library were thawed and recovered
239 overnight in 50 ml SC media lacking uracil. An aliquot of 1 ml (~ 1.8 x 10⁶ cells) was pelleted
240 and frozen as the baseline input sample, and equal aliquots were used to inoculate each of four
241 40 ml cultures of SC media either lacking uracil (nonselective) or lacking both uracil and
242 histidine and optionally containing the competitive inhibitor 3-AT (selective, **Supplementary**
243 **Table 2**). Cultures were maintained at 30°C and checked at 24h, 40h, and 64h. After reaching
244 log-phase (OD 600 >= 0.5), each culture was serially passaged by inoculating 1 ml into 40 ml
245 fresh media.

246 Input and post-selection cultures were pelleted at 16000 x *g* and frozen at -20°C. Gal4
247 plasmids were recovered by spheroplast preparation and alkaline lysis miniprep using the Yeast
248 Plasmid Miniprep II kit as directed (Zymo Research). Two-stage PCR was then used to amplify
249 and prepare sequencing libraries to count the plasmid-tagging tags. In the first step, 2.5ul of
250 miniprep product was used as template in 25ul reactions with KHF RM HS, with primers
251 flanking the tag cassette (“GAL4_BC_AMP_F”/“GAL4_BC_AMP_R”), using the program
252 “GAL4_BARCODE_PCR_ROUND1” for 15-17 cycles. The resulting product was used directly
253 as template (1ul, without cleanup) for the second-stage PCR reaction to add Illumina flowcell-

254 compatible adaptors as well as sample-indexing barcodes to allow pooled sequencing (forward
255 primer “GAL4_ILMN_P5”, and reverse primer one of “SHARED_BC_REV_###”). For the
256 second round, the cycling program “GAL4_BARCODE_PCR_ROUND2” was followed for 5-7
257 cycles. Tag libraries were cleaned up with AmpPure XP beads (2 volumes beads+buffer) and
258 were sequenced across several runs on Illumina MiSeq, GAIx, and HiSeq instruments
259 (**Supplementary Table 8**), using 25-50bp reads.

260 Gal4 enrichment scores

261 Tag reads were demultiplexed to the corresponding sample using a 9 bp index read,
262 allowing for up to two mismatches. Tag reads lacking the proper flanking sequences or
263 containing ambiguous ‘N’ base calls were discarded, and tags were required to exactly match the
264 tag of a single subassembled haplotype. After application of these filters, 18.6% of raw tag reads
265 were discarded. Per-tag histograms were prepared by counting the number of occurrences of
266 each of the remaining tags, and normalizing to account for differing coverage over each library
267 by dividing by the sum of tag counts.

268 We calculated effect scores for each amino acid mutation by summing the read counts of
269 tags corresponding to all the subassembled clones carrying that mutation as a singleton, divided
270 by the equivalent sum for wild-type clones, and taking a log-ratio between the selection and

271 input samples: $e_{MUT\ i} = \log_2 \left(\frac{\sum_{TAG\ j \in MUT\ i} r_{SEL,j} + 1}{\sum_{TAG\ k \in WT} r_{SEL,k} + 1} \right) - \log_2 \left(\frac{\sum_{TAG\ j \in MUT\ i} r_{INPUT,j} + 1}{\sum_{TAG\ k \in WT} r_{INPUT,k} + 1} \right)$, where $r_{SEL,j}$

272 and $r_{INPUT,j}$ are the read counts of tag j in the selected and input samples, respectively.

273 Evolutionarily conserved residues in Zn₂Cys₆ domains were identified by querying
274 HHblits³⁷ with Gal4 residues 1-70, and were displayed using Weblogo³⁸. To compare core and
275 outward-facing residues within the dimerization helix, residues 51-65 were each scored for
276 distance to the overall structure’s solvent-exposed surface predicted using MSMS³⁹ (using the

277 Gal4(1-100) crystal structure, PDB accession 3COQ). Residues with above-median distance to
278 the surface were considered ‘core’, and those with below-median distance were considered
279 ‘exposed’, and the $\log_2 E$ values of the two subsets were compared by the Mann-Whitney U test.

280 **Gal4 effect size validations**

281 For qualitative validation of effect sizes, eight individual alleles (C14Y, K17E, K25W,
282 K25P, L32P, K43P, K45I, and V57M) were re-created by conventional site-directed mutagenesis
283 and assayed for growth defects by a spotting assay (**Supplementary Fig. 7**). These included
284 loss-of-function (C14Y, K17E, and L32P) and hypomorphic alleles (V57M) from initial screens,
285 which conferred growth rates in the spotting assay that agreed with their relative depletion in the
286 deep mutational scan. We likewise validated a novel predicted hypomorphic allele (K25P) and
287 confirmed the slight growth advantage conferred by three alleles from our bulk measurements
288 (K25W, K43P, and K45I). Each allele was individually introduced into p416CYC-Gal4Wt-1-196
289 using the Quickchange mutagenesis kit (Agilent) following the manufacturer’s directions.
290 Mutant colonies were minipreped and verified by capillary sequencing, and transformed into
291 PJ69-4alpha by LiAc treatment. Following transformation, a single yeast colony transformed by
292 mutant or wild-type Gal4 constructs was picked and expanded in overnight culture, and back-
293 diluted to OD 0.2 and allowed to return to mid-log phase before spotting ten-fold dilutions
294 starting with an equal number of cells onto nonselective plates (SC lacking uracil) or selective
295 plates (SC lacking uracil and histidine, supplemented with 5 mM 3-AT).

296

297 **Methods-only References**

- 298 25. Maurer, K. *et al.* *PLoS ONE* **1**, e34 (2006).
299 26. Nakamura, Y., Gojobori, T. & Ikemura, T. *Nucleic Acids Res.* **28**, 292 (2000).
300 27. Mumberg, D., Müller, R. & Funk, M. *Gene* **156**, 119–122 (1995).
301 28. Gill, G. & Ptashne, M. *Nature* **334**, 721–724 (1988).

- 302 29. Li, H. *arXiv* (2013).
- 303 30. Lahr, D. J. G. & Katz, L. A. *BioTechniques* **47**, 857–866 (2009).
- 304 31. Williams, R. *et al. Nature Methods* **3**, 545–550 (2006).
- 305 32. LeProust, E. M. *et al. Nucleic Acids Res.* **38**, 2522–2540 (2010).
- 306 33. Kosuri, S. & Church, G. M. *Nature Methods* **11**, 499–507 (2014).
- 307 34. James, P., Halladay, J. & Craig, E. A. *Genetics* **144**, 1425–1436 (1996).
- 308 35. Gietz, R. D. & Woods, R. A. *Meth. Enzymol.* **350**, 87–96 (2002).
- 309 36. Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R. & Fields, S. *RNA* **19**, 1537–1551
- 310 (2013).
- 311 37. Remmert, M., Biegert, A., Hauser, A. & Söding, J. *Nature Methods* **9**, 173–175 (2012).
- 312 38. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. *Genome Research* **14**, 1188–
- 313 1190 (2004).
- 314 39. Sanner, M. F., Olson, A. J. & Spehner, J. C. *Biopolymers* **38**, 305–320 (1996).