Supplementary Information for

Massively Parallel Single Amino Acid Mutagenesis

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

Online Methods

Mutagenic primer preparation

Mutagenic primers were electrochemically synthesized on a 12,432-feature programmable DNA microarray and released into solution by CustomArrray, Inc[25]. For Gal4 (GI #6325008), codons 2-65 were each replaced with the optimal codon in *S. cerevisae* corresponding to one of the 19 other amino acids[26], a stop codon (TAA), or an in-frame deletion, for a total of 1,344 oligos, each synthesized in duplicate (for a total of 2*64*(19+1+1)=2,688 array features). For p53 (GI #120407068), codons 1-393 were replaced with fully degenerate bases ("NNN") during synthesis, such that primer molecules synthesized within a single spot on the array are degenerate for the triplet corresponding to a single residue, for a total of 393 oligos, each synthesized in triplicate (for a total of 3*393=1,179 array features).

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

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

The resulting oligo pools were further amplified with adaptors modified to contain a deoxyuracil base at the 3' terminus. This second-round amplification was carried out in 50ul reactions, using 1ul of the previous amplification reaction (at a 1:4 dilution in dH_2O) as template, following cycling program "ADO_KR". Each reaction included 25ul Kapa Robust Hot Start ReadyMix (which is not inhibited by uracil-containing templates), amplification primers at 500 nM each ("L_"GAL4DBD"/"R_GAL4DBD_U" or "L_TP53"/"R_TP53_U"), and SYBR Green I at 0.5X. Immediately following PCR, each library was denatured at 95ºC for 30 seconds, and then snap cooled on ice. To cleave the "R" adaptors, 2 U USER enzyme mix (New England Biolabs) was added, and each reaction was incubated for 15 minutes at 37ºC. Finally each reaction was supplemented by 2.5ul of a 10 uM stock of the corresponding "L" primer ("L_GAL4DBD" or "L_TP53"), followed by one final cycle of annealing/priming/extension. Amplification products were purified as before on Zymo columns. Gel electrophoresis confirmed that each resulting library was a mixture of off-product flanked on both sides by adaptors (108 bp), and the desired product with only "L" adaptors (84 bp, **Supplementary Fig. 11**).

Wild-type template preparation

The full-length Gal4 open reading frame was amplified from genomic DNA of *S. cerevisae* strain BY4741 and directionally cloned into the yeast shuttle vector p416CYC, a single-copy CEN plasmid with the CYC1 promoter[27] by digestion with SmaI and ClaI (New England Biolabs), using the InFusion cloning kit (Clontech). Subsequently, an N-terminal

truncation was prepared by amplifying residues 1-196 from the original clone using the primer pairs GAL4_CLONE_F and GAL4_NTERM_R, and recloning into p416CYC to create p416CYC-Gal4Wt-1-196. This fragment retains the same DNA-binding specificity as full-length Gal4 and is sufficient for transcriptional activation[14]. Enforced expression of full-length Gal4 causes cellular toxicity by aberrantly sequestering the transcriptional machinery, in an effect called squelching[28]. A similar effect is observed for Gal4 1-196 (i.e., loss-of-function alleles are more fit than wild-type ones under non-selective growth, **Supplementary Fig. 6**) but to a much lesser degree than for the full length protein. For p53, a wild-type clone with a C-terminal GFP fusion was purchased from OriGene (#RG200003).

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

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

Mutagenic primer extension

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

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

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

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

PALS library cloning

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

randomized 16mer ("P416CYC_BC_CAS") was then PCR-amplified using primers "P416CYC_AMP_BC_CAS_F"/"P416CYC_AMP_BC_CAS_R" and cycling program "MAKE_BC_CAS", to add priming sites for later tag counting during Gal4 functional selections, and to add flanking AgeI and MfeI sites. The resulting tag cassette amplicon was directionally cloned into the modified p416CYC vector by double-digestion with AgeI-HF and MfeI-HF (NEB) and transformed into ElectroMax DH10B electrocompetent *E. coli* (Invitrogen), to yield ~$9.2x10^6$ distinctly tagged clones. The resulting library, p416CYC-bc, was expanded by bulk outgrowth and purified by midiprep using the ChargeSwitch Pro Midi kit (Invitrogen). Next, 15 ug of p416CYC-bc was digested with 40U SmaI (NEB) for 1 hr at 25ºC in 60ul, followed by addition of 20U ClaI (NEB), digestion for 1 hr at 37ºC, and purification by MinElute column (Qiagen). To insert the Gal4 DBD PALS library, 50 ng of the final PALS PCR product was combined with 10ng SmaI/ClaI linearized p416CYC-bc vector and directionally cloned using the InFusion HD kit (ClonTech), as directed. Libraries were transformed by electroporation into 10-beta electrocompetent *E. coli* (NEB), and bulk transformation cultures were expanded overnight in 25ml LB + ampicillin (50 ug/ml) at 37ºC, shaking at 250 rpm. Due to the large number of vector copies present in the cloning reaction, pairing of Gal4 mutant inserts with tag is essentially sampling with replacement; the number of positive clones (~$9.0x10^5$) is less than the number of tags by approximately an order of magnitude, so only ~0.45% of tags are expected to be paired with two different inserts.

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

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

Clone subassembly sequencing

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

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

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

Tag-directed clone subassembly

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

https://code.google.com/p/cutadapt/), and read pairs without recognizable tag-flanking adaptors were excluded from further analysis. Insert-end reads were aligned to the Gal4 or p53 wild-type clone sequence using bwa mem[29] (with arguments "-z 1 -M"), and alignments were sorted and grouped by their corresponding clone tag. To properly align the programmed in-frame codon deletions included in the Gal4 PALS library, bwa alignments were realigned using a custom implementation of Needleman-Wunsch global alignment with a reduced gap opening penalty at codon start positions (match score=1, mismatch score=-1, gap open in coding frame =-2, gap open elsewhere=-3, gap extend=-1). A consensus haplotype sequence was determined for each tag-defined read group by incorporating variants present in the group's aligned reads at sufficient depth. Spurious mutations created by sequencing errors, or mutations present at low allele frequency arising from linking two haplotypes to the same tag were flagged and discarded by requiring the major allele at each position (either wild-type or mutant) to be present with a frequency of ≥80%, ≥75% and ≥66%, for read depths ≥20, 10-19, or 3-10, respectively, considering only bases with quality score ≥20. Tag groups with fewer than three reads (Gal4 DBD) or 20 reads (p53) were discarded, as were groups not meeting the major allele frequency threshold across the entire target (Gal4 DBD) or a minimum of 1 kbp (p53). Consensus haplotypes were validated by Sanger sequencing of individual colonies from each tagged plasmid library (**Supplementary Fig. 12** and **Supplementary Table 1**).

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

Chimerism is a technical challenge commonly encountered while amplifying libraries of homologous sequences[30], when incomplete strand extension products in one cycle of amplification act as primers in the subsequent cycle. Future optimization efforts will be directed at quantifying and mitigating this phenomenon by manipulating input template concentration and minimizing amplification cycles, or alternatively using droplet PCR[31]. To reduce the impact of synthesis errors, PALS uses short oligonucleotides (90 nt), but it will nevertheless benefit from ongoing developments in high-fidelity synthesis[32]. In addition, as single-base deletions are the dominant synthesis error mode[33], stringently size-selecting primer libraries may further enrich for primers lacking undesirable secondary mutations. Another strategy would fuse libraries in-frame to a selectable marker in the bacterial cloning host, although our preliminary observations suggest that such selection is inefficient for proteins that do not fold or express well in E. coli. For p53, because codon substitutions were encoded as "NNN", the origin of secondary mutations could not be distinguished between synthesis errors, PCR errors, or chimerism between fragments each bearing a single codon swaps.

Although PALS is intended to create single-mutant clones, for applications such as protein engineering, it may be useful to obtain multiple mutations per copy. This could be accommodated by applying PALS serially, to first create a library of single-mutant copies which would then be used as the starting template for the second round. In the context of typical-length genes, the multi-mutation space is so large (e.g., for TP53 double mutants, choose(393,2)*19*19=2.78x10^7 possibilities) that it may be technically impractical to construct, much less survey, the entire space. By serially applying PALS, however, it could be possible to focus on a defined subspace using a subset of mutagenic primers in either or both rounds.

Gal4 functional selections

Gal4 DBD PALS libraries were transformed into chemically competent *S. cerevisae* strain PJ69-4alpha[34] prepared using a modified LiAc-PEG protocol, as previously described[35,36]. After transformation, cells were allowed to recover for 80 minutes at 30ºC shaking at 250 rpm. To select for transformants, cultures were spun down at 2000 x *g* for 3 min, resuspended and grown overnight at 30ºC in 40 ml SC media lacking uracil. Plating 0.25% of the recovery culture prior to outgrowth indicated a library of ~ $2x10^5$ transformants. Following overnight outgrowth, glycerol stocks were prepared from the transformation culture and stored at -80ºC.

Frozen stocks of yeast carrying the Gal4 DBD PALS library were thawed and recovered overnight in 50 ml SC media lacking uracil. An aliquot of 1 ml (~ 1.8 x 10^6 cells) was pelleted and frozen as the baseline input sample, and equal aliquots were used to inoculate each of four 40 ml cultures of SC media either lacking uracil (nonselective) or lacking both uracil and histidine and optionally containing the competitive inhibitor 3-AT (selective, **Supplementary Table 2**). Cultures were maintained at 30ºC and checked at 24h, 40h, and 64h. After reaching log-phase (OD 600 >= 0.5), each culture was serially passaged by inoculating 1 ml into 40 ml fresh media.

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

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

Gal4 enrichment scores

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

We calculated effect scores for each amino acid mutation by summing the read counts of tags corresponding to all the subassembled clones carrying that mutation as a singleton, divided by the equivalent sum for wild-type clones, and taking a log-ratio between the selection and input samples: $e_{\mathrm{MUT}\ i} = \log_2\left(\frac{\sum_{\mathrm{TAG}\ j \in \mathrm{MUT}\ i}\ r_{\mathrm{SEL},j}+1}{\sum_{\mathrm{TAG}\ k \in \mathrm{WT}}\ r_{\mathrm{SEL},k}+1}\right) - \log_2\left(\frac{\sum_{\mathrm{TAG}\ j \in \mathrm{MUT}\ i}\ r_{\mathrm{INPUT},j}+1}{\sum_{\mathrm{TAG}\ k \in \mathrm{WT}}\ r_{\mathrm{INPUT},k}+1}\right)$, where $r_{\mathrm{SEL},j}$ and $r_{\mathrm{INPUT},j}$ are the read counts of tag j in the selected and input samples, respectively.

Evolutionarily conserved residues in Zn_2Cys_6 domains were identified by querying HHblits[37] with Gal4 residues 1-70, and were displayed using Weblogo[38]. To compare core and outward-facing residues within the dimerization helix, residues 51-65 were each scored for distance to the overall structure's solvent-exposed surface predicted using MSMS[39] (using the

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

Gal4 effect size validations

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

Methods-only References

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

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