Here we describe a C-SWAT library for high-throughput tagging of *Saccharomyces cerevisiae* open reading frames (ORFs). In 5,661 strains, we inserted an acceptor module after each ORF that can be efficiently replaced with tags or regulatory elements. We validated the library with targeted sequencing and tagged the proteome with bright fluorescent proteins to quantify the effect of heterologous transcription terminators on protein expression and to localize previously undetected proteins.

Genome-wide libraries of strains in which every ORF is fused to a constant tag are valuable resources for proteome-wide studies in *S. cerevisiae*1–5. However, the construction of such libraries is costly and time-consuming, which hampers genome-wide endeavors with novel tags.

To overcome these limitations, we recently developed the SWAT-Tag (SWAT) approach for high-throughput tagging of yeast ORFs and used it to N-terminally tag proteins of the endomembrane system6. This approach requires the one-time construction of SWAT strains in which individual ORFs are marked with an acceptor module (Fig. 1a). New strains can be rapidly derived from SWAT strains through the use of automated procedures (detailed below) to swap the acceptor module for practically any tag or regulatory element provided on a donor plasmid (Fig. 1a).

Here we developed a genome-wide C-SWAT (C-terminal SWAT-Tag) library. This arrayed library enables genome engineering at 3′ ends of yeast ORFs and can be used for high-throughput C-terminal protein tagging. We used conventional PCR targeting7,8 to insert a C-SWAT acceptor module before the stop codon of individual ORFs at endogenous chromosomal loci (Fig. 1a). The acceptor module consists of homology arms (L3 and L4, for subsequent recombination) and the acceptor module before the stop codon of individual ORFs (Fig. 1a). New strains can be rapidly derived from SWAT strains through the use of automated procedures (detailed below) to swap the acceptor module for practically any tag or regulatory element provided on a donor plasmid (Fig. 1a).

For strain validation in high throughput, we developed a targeted sequencing approach (Anchor-Seq) to sequence the junctions between the 3′ end of each tagged ORF and the 5′ end of the acceptor module (Fig. 1b). In Anchor-Seq, genomic DNA is isolated from a pooled library of strains, where a different ORF is modified in each strain. The junctions of interest are then selectively amplified using Vectorette PCR and sequenced (Fig. 1b, Supplementary Fig. 1 and Methods). Vectorette PCR relies on Vectorettes, oligonucleotides containing a central mismatch region, to selectively amplify DNA regions adjacent to a known sequence9 (Supplementary Fig. 1). We performed Anchor-Seq on pools of six replicates of the C-SWAT library, corresponding to six independent transformants for each ORF. In total, we obtained validated C-SWAT strains for 93% of verified or uncharacterized *S. cerevisiae* ORFs and for 158 dubious ORFs (Fig. 1c and Supplementary Table 1). The result was an arrayed C-SWAT library composed of 5,661 strains.

To tag ORFs with the C-SWAT library, we introduced a construct for conditional expression of the I-SceI endonuclease and a donor plasmid carrying the desired tag into C-SWAT strains in high throughput by automated genetic crossing with a donor strain (Fig. 1a, Methods, and Supplementary Note 1). Three types of donor plasmids with different selection strategies can be used: type I for markerless replacement of the acceptor module with just the tag sequence, type II for selection of tagging events via reconstitution of the hygromycin resistance marker (hph), and type III for selection of tagging events with a new selection marker (Supplementary Fig. 2a and Supplementary Note 1). Using C-SWAT strains for 20 high-expression genes as a test case, we observed average tag swapping efficiency of ~98% with a type I donor and >99% with the other two donor types (Supplementary Fig. 2b). This demonstrates that the C-SWAT library can be used for high-throughput strain construction without the need for subsequent clonal selection. Endogenous repetitive sequences surrounding the tag integration site potentially interfere with markerless tag swapping in 1–4% of ORFs10. We note that after tag swapping, the L3 sequence becomes a linker between the ORF and the tag, irrespective of the type of donor plasmid used (Fig. 1a and Supplementary Fig. 2a). Moreover, with markerless tagging, the L4 sequence is part of the transcript, located between the tag and the endogenous 3′ untranslated region (UTR) (Supplementary Fig. 2a). The L4 sequence used in the C-SWAT library is neutral, as it does not interfere with expression of SPS100 and UGA2, two genes that are strongly regulated by their 3′ UTRs (ref. 11 and data.
We then quantified protein expression levels in these libraries (Supplementary Fig. 5a). mNeonGreen-tagged YPR170w-B localized to the endoplasmic reticulum, whereas Ypr170w-A localized to the vacuolar membrane (Fig. 2f and Supplementary Fig. 5a). This suggests different functions for the two proteins, although we cannot exclude the possibility that tagged Ypr169w-a is expressed only because the tag disrupts the splicing and expression of YPR170W-B.

With updates to the yeast genome annotation, 80 ORFs included in the C-SWAT library have been reclassified from dubious to verified or uncharacterized since we began constructing the library (Supplementary Table 4). We note that the remaining dubious ORFs potentially also encode functional proteins, as their distribution of expression levels is similar to that for the reclassified ORFs in the mNG-I library (Supplementary Fig. 5b).

With the C-SWAT library, the yeast ORFeome can be efficiently manipulated to generate libraries with a variety of tags for protein or RNA detection, to study regulation of gene expression, or to explore genomic position effects. Although the experiments described herein were performed in arrayed format, the C-SWAT and derived libraries could in principle be used for pooled experiments in which the unique ORF-tag junctions serve as barcodes (Supplementary Note 2). We hope that the simplicity and cost-effectiveness of C-SWAT will make the construction of custom genome-wide libraries routine and facilitate systematic studies.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41592-018-0045-8.

Received: 28 November 2017; Accepted: 4 May 2018; Published online: 9 July 2018

![Diagram](https://example.com/diagram.png)

**Fig. 1** | Design, construction, and validation of the C-SWAT library. 
**a** Outline of the C-SWAT approach. A C-SWAT acceptor module is inserted into the genome before the stop codon of yeast ORFs. A construct for conditional expression of the I-SceI endonuclease (not shown) and a donor plasmid carrying the desired tag are then introduced into the C-SWAT strains by transformation or genetic crossing. After expression, I-SceI induces double-strand breaks (DSBs) at the indicated positions (● in the acceptor module and the donor plasmid. DSB repair by homologous recombination leads to replacement of the acceptor module with the tag. **b** Outline of the Anchor-Seq targeted sequencing approach. A library of strains with different ORFs (magenta) modified with a constant tag (green, e.g., C-SWAT acceptor module) is pooled and subjected to Anchor-Seq (Supplementary Fig. 1). **c** Composition of the C-SWAT library. ORFs are classified as verified, uncharacterized, or dubious (i.e., unlikely to encode functional proteins). 6,014 ORFs were selected for tagging with the C-SWAT acceptor module (selected). C-SWAT strains validated with Anchor-Seq were obtained for 5,661 ORFs (tagged).
Fig. 2 | High-throughput protein tagging with the C-SWAT library. **a**, Donor plasmids for tagging the yeast proteome with mNeonGreen and mScarlet-I fluorescent proteins using the C-SWAT library. Symbols are defined as in Fig. 1a and Supplementary Fig. 2a. **b–e**, Analysis of protein expression with mNeonGreen and mScarlet-I libraries. Shown are fluorescence measurements of colonies (median of three technical replicates). Background fluorescence was determined from control strains not expressing a fluorescent protein. **b**, Number of protein fusions detected in each library as a function of a fluorescence threshold. Strain fluorescence is expressed in units of background fluorescence. 4,312, 4,537, and 4,301 strains in the mNG-I, mNG-II, and mSC-II libraries, respectively, had a fluorescence signal at least 1.2-fold above background (vertical red dashed line). The number of protein fusions detected in the GFP library with fluorescence microscopy\(^{20}\) is shown for comparison. **c**, Distribution of differences in protein expression levels (background-corrected strain fluorescence) between mNG-I and mNG-II libraries (median = 0.83; red dashed line). **d**, Correlation of endogenous transcription terminator activity for each ORF, measured in ref. \(^{10}\), and differences in protein expression levels between mNG-I and mNG-II strains. Spearman’s rank correlation coefficient \(r = 0.509\). **e**, Fluorescence levels of mNG-I strains (in units of background fluorescence) expressing 207 previously undetected proteins compared with the entire mNG-I library. Only strains with fluorescence at least 1.2-fold above background were considered. **f,g**, Fluorescence microscopy of strains from mNG-I and mNG-II libraries expressing 207 previously undetected proteins tagged with mNeonGreen. **f**, Examples of fusions with different subcellular localizations. The two images for each protein were acquired and processed identically. Scale bar, 5\(\mu\)m. **g**, Summary of observed subcellular localizations from one experiment. ER, endoplasmic reticulum; cyto, cytosol; nucl, nucleus or nuclear periphery; mito, mitochondria; vac, vacuole; punct, punctate; peri, cell periphery; ambig, ambiguous; ND, expression not detected.

**References**
1. Huh, W.-K. et al. *Nature* **425**, 686–691 (2003).
2. Ghaemmaghami, S. et al. *Nature* **425**, 737–741 (2003).
3. Gavin, A.-C. et al. *Nature* **440**, 631–636 (2006).
4. Tarassov, K. et al. *Science* **320**, 1465–1470 (2008).
5. Khmelinskii, A. et al. *Nature* **516**, 410–413 (2014).
6. Yaffe, I. et al. *Nat. Methods* **13**, 371–378 (2016).
7. Baudin, A., Özer-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. *Nucleic Acids Res.* **21**, 3329–3330 (1993).
8. Wach, A., Brachat, A., Alberti-Segui, C., Rebischung, C. & Philippsen, P. *Yeast* **13**, 1065–1075 (1997).
9. Arnold, C. & Hodgson, I. J. *PCR Methods Appl.* **1**, 39–42 (1991).
10. Khmelinskii, A., Meurer, M., Duishoev, N., Delhomme, N. & Knop, M. *PLoS One* **6**, e23794 (2011).
11. Bunina, D. et al. *Nucleic Acids Res.* **45**, 11144–11158 (2017).
12. Tsien, R. Y. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
13. Shaner, N. C. et al. *Nat. Methods* **10**, 407–409 (2013).
14. Bindels, D. S. et al. *Nat. Methods* **14**, 53–56 (2017).
15. Ho, B., Barryhukova, A. & Brown, G. W. *Cell Syst.* **6**, 192–205 (2018).
16. Yamanishi, M. et al. *ACS Synth. Biol.* **2**, 337–347 (2013).
17. Longtine, M. S. et al. *Yeast* **14**, 953–961 (1998).
18. Janke, C. et al. *Yeast* **21**, 947–962 (2004).
19. Cherry, J. M. et al. *Nucleic Acids Res.* **40**, D700–D705 (2012).
20. Cheong, Y. T. et al. *Cell* **161**, 1413–1424 (2015).

**Acknowledgements**
We thank the CellNetworks Deep Sequencing Core Facility (Heidelberg University) and the Genomics Core Facility (EMBL), and acknowledge generous support from the DFG for data storage (LD5F2). This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Center SFB1036 (M.K., A.K., T.P.D., and M.K.L.), the Weizmann Institute of Science (E.S. and E.D.L.), the China Scholarship Council (Y.D.), fellowships from the HBIGS graduate school (I.K., K.H., and E.H.), an SFB1036 travel grant (E.S.), and the Alexander von Humboldt Foundation (M.S.), and partially by DFG grant KN498/11-1 (M.K.), Israel Science Foundation grants 1775/12 and 2179/14 (E.D.L.), and HFSP Career Development Award CDA00077/2015 (E.D.L.).

**Author contributions**
M.K., M.M., E.D.L., and A.K. planned the work. Y.D. and M.M., together with E.S., constructed the library, with help from B.C.B., Y.D., K.H., E.H., D.K., I.K., M.S., K.V.L., A.K., and M.K.E.S. and E.D.L. developed Anchor-Seq and E.D.L. analyzed the sequencing data, with input from M.M. and K.H. Y.D., M.M., I.K., A.K., and D.K. generated and analyzed the mNeonGreen and mScarlet-I libraries. A.K. and M.K. wrote the manuscript with E.D.L. and E.S., with input from all other authors.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0045-8.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.K. or E.D.L. or M.K.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
**Methods**

**Construction of the C-SWAT library. Acceptor module.** The acceptor module used to construct the C-SWAT library (plasmid pMaM471; Supplementary Table 5) is composed of the following elements:

- linker L3 (5'-GGAGTGTCAGATGGGGTTATAGATGGGACG-3'), which contains the S3 primer annealing site (underlined) for gene tagging by PCR targeting
- STOP codon
- terminator sequence of the CYC1 gene from Saccharomyces paradoxus
- the recognition sequence for the I-SceI endonuclease
- the URA3 gene with its endogenous promoter and terminator from S. cerevisiae
- hphΔN sequence coding for a C-terminal fragment (amino acids 146–342) of the hph (hygromycin-resistance-encoding gene) marker
- terminator sequence of the ALG9 gene from S. paradoxus
- linker L4 (5'-AGTCTTCTGGATGCTGGACGAGATGGATTGGTGGAT-3'), which contains the S2 primer annealing site (underlined) for gene tagging by PCR targeting

The L4 linker was previously used in markerless tagging constructs that did not disrupt antiseNSE transcription13, which shows that this sequence is unlikely to interfere with regulatory elements downstream of ORFs.

**Library background strain.** The strain BY47411 (Supplementary Table 6), used to construct the most popular collections of yeast strains such as the knockout and GFP libraries14,15, was chosen as the background strain for the C-SWAT library. All additional elements required for the SWAT procedure (a donor plasmid and a construct for inducible expression of the I-SceI endonuclease) can be introduced into C-SWAT acceptor strains by direct transformation or by genetic crossing with the donor strain YMaM639 (Supplementary Table 6) carrying the desired donor plasmid.

**Donor strain and plasmids.** The donor strain YMaM639 was constructed using the strain Y28395 (Supplementary Table 6). In this strain the leu2Δ locus carries the GAL1pr-NLS-I-SCG1-natNT2 construct for galactose-inducible expression of the I-SceI endonuclease with a nuclear localization signal, which increases the efficiency of tag swapping. In addition, this strain also contains the can1Δ:STE2pr-SPHIS5 and hpyLP1:STE3pr-LEU2 markers for selection of MATa or MATa alpha haploids at the end of the automated genetic crossing procedure via synthetic genetic array (SGA) methodology15.

Three different template plasmids were used to construct donor plasmids (Supplementary Table 5):

- pMaM482: type I donor template (for markerless tag swap)
- pMaM484: type II donor template (for tag swap with reconstitution of the hph marker)
- pMaM496: type III donor template (for tag swap with introduction of the kanMX6 marker)

The backbone of all donors is pRS41K. For donor type III, the linker L2 (5'-GCGGCTTAAGGATCCGTGCTCGAG-3') was added in between the S2 primer annealing site (underlined) for gene tagging by PCR targeting

**Selection of ORFs and primer design.** We selected 6,071 yeast ORFs for tagging with the acceptor module: all 5,797 verified or uncharacterized ORFs and 274 dubious ORFs that do not overlap with any verified or uncharacterized ORFs with the acceptor module: all 5,797 verified or uncharacterized ORFs and 274 dubious ORFs that do not overlap with any verified or uncharacterized ORFs. For each ORF, S2/S3 primers for PCR amplification of the tagging module were designed as previously described15. For 4,081 ORFs, primers synthesized on the basis of the yeast genome sequence from December 2009 (Saccharomyces Genome Database) were available from a previous study15. For the remaining 1,990 ORFs, primers were designed on the basis of the yeast genome sequence from July 2016. In this second set, ORFs with identical S2 and S3 primer sequences were identified (e.g., HXT15 and HXT16). For such cases, only one set of primers was synthesized and assigned to the first ORF by alphabetical order of systematic names. This reduced the set of 1,990 ORFs to 1,933. The 6,014 pairs of S2/S3 primers (Supplementary Table 1) were obtained from IDT (Integrated DNA Technologies) in 96-well format, such that each well contained a mixture of S2/S3 primers for a different ORF at 5µM concentration.

**Strain construction.** The acceptor module was amplified by PCR in 96-well format using plasmid pMaM471 as template and ORF-specific S2/S3 primers in each well as follows. Cooled 96-well PCR plates (Agriculture, 4-ti-0990) were filled with 40 µl per well of a PCR mix using a Liquidator 96 channel manual pipettor (Mettler Toledo):
The abundance (cycle threshold (CT) values) of on-target and off-target sequences, before and after selective amplification, were determined by StepOnePlus Software v2.3 (Thermo Fisher Scientific). Enrichment was defined as the difference of CT values (ΔCT) and calculated as 2^{ΔCT}, where Δ is the ΔCT between off- and on-target sequences before selective amplification, and B is the ΔCT for these sequences after selective amplification. We observed enrichments typically in the ~10^{-7}~10^{-10}-fold range.

Samples were normalized to 10 nM final concentration and subjected to sequencing (Illumina MiSeq PE300, V3 flowcell). All six libraries were sequenced simultaneously. For demultiplexing, the barcodes added during the second PCR step matched as well as six possible frameshifts (−3, −2, −1, +1, +2, +3), and we denote the corresponding counts as M_{i} (first half) and X_{i} (second half). We used three criteria to identify correct clones:

1. the number of strict matches had to be above 5,
2. the number of strict matches had to exceed the total number of erroneous reads, and
3. the number of partial matches could not exceed strict matches by more than fivefold.

That is, ORF \( i \) was considered correctly tagged if \( M_{i} > 5, M_{i} > (F_{i}^− + F_{i}^+), F_{i}^{−1}, F_{i}^{−2}, F_{i}^{−3}, F_{i}^{+1}, F_{i}^{+2}, F_{i}^{+3}), 5M_{i} > X_{i}, 5M_{i} > X_{2} \). In total we identified at least one correct clone for 5,661 ORFs (Fig. 1c and Supplementary Table 1). These clones, one for each ORF, were arranged in 96-well plates forming the C-SW library v1.0.

Tag swap with the C-SW library. Donor plasmids pmAM482, pmAM484, pYD10, pYD11, pYD13, and pYD14 (Supplementary Table 5) were transformed into the YMaH639 donor strain.

To test the efficiency of tag swapping (Supplementary Fig. 2b), we selected 20 C-SW strains for highly expressed proteins to facilitate single-cell fluorescence measurements with flow cytometry. The strains were randomly selected from the first plate of the C-SW library, which mostly contains C-SW strains for highly expressed proteins. Strains with tagged ribosomal subunits or histones or corresponding to proteins expressed at less than 20,000 molecules per cell were excluded. This set of 20 strains was crossed with four donor strains carrying different donor plasmids: YYD2 (type I mNeonGreen donor), YYD3 (type II mNeonGreen donor), YYD8 (type III mNeonGreen donor), and YYD6 (empty type II donor) (Supplementary Table 6).

The full C-SW library was crossed with the following four donor strains: YYD2 (type I mNeonGreen donor), YYD3 (type II mNeonGreen donor), YYD4 (type II mScarlet-I donor), and YYD5 (empty type I donor) (Supplementary Table 6). We carried out crossing and subsequent tag swapping by sequentially pinning the strains on appropriate media using a ROTOR HDA pinning robot (Singer Instruments) in 1,536-column format according to the synthetic genetic array (SGA) procedure as follows:

- mating of C-SW and donor strains on YPD plates (10 g/1 yeast extract (BD Biosciences, 212750), 20 g/1 peptone (BD Biosciences, 211677), 20 g/1 glucose (Merck, 108337), 20 g/1 agar (BD Biosciences, 214010)), 1 d at 30°C
- selection of diploids on SC-(MSG)-Ura + G+4-18 plates (1.7 g/1 yeast nitrogen base without amino acids and ammonium sulfate (BD Biosciences, 233520), 1 g/1 monosodium glutamate (MSG) (Sigma-Aldrich, G1626), 2 g/1 amino acid mix SC-(MSG)-Ura (glutamic acid replaced by MSG), G + 418 (200 mg/1, Biochrom, A291–25), 20 g/1 glucose, 20 g/1 agar), 1 d at 30°C
- sporulation on SPO plates (20 g/1 potassium acetate (Sigma-Aldrich, 25059), 20 g/1 agar), 5 d at 32°C
- selection of haploids, step 1, on SC-(MSG)-Ura/His/Arg/Lys + canavanine/thioly-sine plates (50 mg/1 canavanine (Sigma-Aldrich, C1625), 50 mg/1 thioreline (Sigma-Aldrich, A2636)), 2 d at 30°C
- selection of haploids, step 2, on SC-(MSG)-Ura/His/Arg/Lys + canavanine/thioly-sine/G-418 plates (50 mg/1 canavanine, 50 mg/1 thioreline, 200 mg/1 G-418), 2 d at 30°C
- selection of haploids, step 3, on SC-(MSG)-Ura/His/Arg/Lys + canavanine/thioly-sine/G-418/clonNAT plates (50 mg/1 canavanine, 50 mg/1 thioreline, 200 mg/1 G-418, 100 mg/1 clonNAT (Werner BioAgents, 5.01)), 2 d at 30°C
- induction of tag swapping on SC-His Gal/Raf plates (6.7 g/1 yeast nitrogen base without amino acids (BD Biosciences, 291940), 2 g/1 amino acid mix SC-His, 20 g/1 galactose (Serva, 22202), 20 g/1 raffinose (Sigma-Aldrich, R0250), 20 g/1 agar), 2 d at 30°C (done twice)
- selection against the acceptor module on SC-His + 5-FOA plates (6.7 g/1 yeast nitrogen base without amino acids, 2 g/1 amino acid mix SC-His, 20 g/1 glucose, 1 g/1 5-FOA (Apollo Scientific, PC4054)), 2 d at 30°C

Finally, strains resulting from the swap of the full C-SW library were pinned on SC-His and grown for 1 d at 30°C before fluorescence measurements of colonies.

Strains swapped to test the efficiency of tag swapping were pinned from SC-His + 5-FOA plates either to (SCMSG)-His + G-418 plates for all three donor types or to (SCMSG)-His + hygromycin plates (200 mg/1 Hygromycin B Gold, Invivogen, anti-hg-3) for the type II donor or to (SCMSG)-His + G-418 plates (200 mg/1 G-418) for the type III donor. Finally, all strains were pinned on SC-His plates and grown for 1 d at 30°C before fluorescence measurements with flow cytometry.

We note that a pinning robot is not essential for tag swapping with the C-SW library. Instead, manual pin tools can be used, as previously described\(^8\), to arrange the strains in 384-colony format and to perform the crossing and subsequent tag swapping.

Flow cytometry. Strains were grown to saturation in 96-well plates (150 μ l of SC-His medium per well) at 30°C, diluted into fresh SC-His medium, and grown for 8 h at 30°C to 2 × 10^7 cells/ml. Fluorescence measurements were performed on a BD FACS Canto RUO (BD Biosciences) equipped with a high-throughput sampler loader, a 488-nm laser and a combination of 505-nm long pass filter and 530/30-nm band pass filter for mNeonGreen detection. Populations were gated for single cells in the G1 phase of the cell cycle using the first peak in the side scatter width (SSC-W) histogram, and 20,000 cells were measured for each strain.

Colonies fluorescence measurements. Strains resulting from the swap of the full C-SW library with mNeonGreen and mScarlet-I donors in three technical replicates were arranged next to each other and three technical replicates of a negative control (tag swap using the empty donor plasmid pmAM482). Strains were pinned on SC-His agar plates with a ROTOR HDA pinning robot (Singer Instruments) and grown at 30°C for 24 h. Fluorescence measurements were performed at 30°C with Infinite M1000 or Infinite M1000 Pro plate readers (Tecan) equipped with stackers for automated plate loading and custom temperature control chambers. Detector gain was set manually to avoid saturation, and measurements were performed at 400-Hz frequency with the flash lamp, with ten flashes averaged for each measurement, in two channels: mScarlet-I (569/10-nm excitation, 593/10-nm emission) and mNeonGreen (506/5-nm excitation, 517/5-nm emission). Measurements were filtered for potentially failed crosses on the basis of colony size after haploid selection. Colony area measurements for each individual plate were median-centered before calculation of median colony size for the entire dataset. Scanned median absolute deviation (MAD) served as a robust estimate of s.d., and colonies within the 0.5th percentile of a normal distribution were used as background. Background values for each strain with a spread of scaled MAD were defined as failed crosses. Tag swaps with less than two successfully crossed replicates were removed from the analysis. Fluorescence intensities for each plate were normalized to the median fluorescence of a reference strain set that was present on every plate. Intensities of sample colonies were either corrected for background by subtraction of the average intensity of negative control colonies or expressed in background units, that is, divided by the average intensity of negative control colonies (Supplementary Table 2). Both background correction and background normalization were performed locally, that is, using the three negative control colonies closest to each sample (also represented by three colonies arranged next to each other). A fluorescence threshold of 1.2, above which a tagged protein was considered to be expressed at detectable levels, was chosen according to distributions of locally normalized fluorescence intensities of negative control colonies. Less than 0.5% of negative control colonies fell above this threshold (Supplementary Fig. 4a), whereas 17.6%, 13.1%, and 15.7% of strains in the mNG-I, mNG-II, and mSC-II libraries, respectively, did not exceed the threshold.

Fluorescence microscopy. Strains were inoculated in 96-well plates in synthetic complete (SC) low-fluorescence medium (SC medium prepared with yeast nitrogen base lacking folic acid and riboflavin\(^9\)) and grown at 30°C to mid-log phase for 7–8 h. 150 μ l of each culture were used for microscopy in glass-bottom 96-well plates (MGB096-1-2-LG-1; Matrical) coated with concanavalin A, as described\(^10\). Imaging was performed on a Nikon Ti-E widefield epifluorescence microscope with a 60× NPLA objective (1.49 NA from Nikon; Nikon), an LED light engine (SpectraX, Lumenz), an sCMOS (scientific complementary metal-oxide semiconductor) camera (Flash4, Hamamatsu) and an autofocus system (Perfect Focus System, Nikon) with either bright field or 469/35-
nm excitation filter (Semrock) and 525/50-nm emission filter (Chroma). Z-stacks of 11 planes with 0.5-µm spacing were recorded with 100-ms exposure time for all strains. Single plane images (Fig. 2f and Supplementary Fig. 5a) and maximum intensity z-projections (Supplementary Table 3 and Supplementary Data) are shown. Subcellular localizations were identified and scored visually.

Availability of resources. The C-SWAT library, the derived mNG-I, mNG-II and mSC-II libraries, all reagents necessary to use the C-SWAT library for high-throughput strain construction, and custom scripts for analysis of Anchor-Seq data are available from the corresponding authors on request.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data that support the findings of this study are available from the corresponding authors on request.

References
21. Huber, F. et al. Cell Rep. 15, 2625–2636 (2016).
22. Brachmann, C. B. et al. Yeast 14, 115–132 (1998).
23. Winzeler, E. A. et al. Science 285, 901–906 (1999).
24. Tong, A. H. Y. & Boone, C. Methods Microbiol. 36, 369–386, 706–707 (2007).
25. Taxis, C. & Knop, M. Biotechniques 40, 73–78 (2006).
26. Knop, M. et al. Yeast 15, 963–972 (1999).
27. Baryshnikova, A. et al. Methods Enzymol. 470, 145–179 (2010).
28. Sheff, M. A. & Thorn, K. S. Yeast 21, 661–670 (2004).
29. Khmelinskii, A. & Knop, M. Methods Mol. Biol. 1174, 195–210 (2014).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

### Experimental design

1. **Sample size**
   - Describe how sample size was determined.
   - **does not apply**

2. **Data exclusions**
   - Describe any data exclusions.
   - **no data was excluded**

3. **Replication**
   - Describe the measures taken to verify the reproducibility of the experimental findings.
   - **all procedures were validated for reproducibility**

4. **Randomization**
   - Describe how samples/organisms/participants were allocated into experimental groups.
   - **does not apply**

5. **Blinding**
   - Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   - **does not apply**

   Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. **Statistical parameters**
   - For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   **n/a** **Confirmed**
   - [ ] The exact sample size \( n \) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - [ ] A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   - [ ] A statement indicating how many times each experiment was replicated
   - [ ] The statistical test(s) used and whether they are one- or two-sided
   - [ ] Only common tests should be described solely by name; describe more complex techniques in the Methods section.
   - [ ] A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   - [ ] Test values indicating whether an effect is present
     - Provide confidence intervals or give results of significance tests (e.g. \( P \) values) as exact values whenever appropriate and with effect sizes noted.
   - [ ] A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   - [ ] Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

   See the web collection on statistics for biologists for further resources and guidance.
7. Software

Describe the software used to analyze the data in this study.

Custom Perl and R scripts, Bowtie2 was used for alignment, StepOnePlus v2.3 (Thermo Fischer Scientific).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

We will make sure that the main resource of this paper (the C-SWAT library) is fully accessible - in fact we have already handed it out to several partners and we are well equipped to ship more copies in an efficient manner and free of charge.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

no antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Yeast strains BY4741, Y7092 and Y8205 were obtained from the laboratory of Charles Boone, Donnelly Center, Toronto

b. Describe the method of cell line authentication used.

marker testing and sequencing

c. Report whether the cell lines were tested for mycoplasma contamination.

no mammalian cell lines were used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

does not apply

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

no animals used

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

no humans used