

# Genome-wide C-SWAT library for high-throughput yeast genome tagging

Matthias Meurer<sup>1,8</sup>, Yuanqiang Duan<sup>1,8</sup>, Ehud Sass<sup>2,8</sup>, Ilia Kats<sup>1</sup>, Konrad Herbst<sup>1</sup>, Benjamin C. Buchmuller<sup>1</sup>, Verena Dederer<sup>1</sup>, Florian Huber<sup>1,5</sup>, Daniel Kirrmaier<sup>3</sup>, Martin Štefl<sup>1,6</sup>, Koen Van Laer<sup>4</sup>, Tobias P. Dick<sup>4</sup>, Marius K. Lemberg<sup>1</sup>, Anton Khmelinskii<sup>1,7\*</sup>, Emmanuel D. Levy<sup>1,2\*</sup> and Michael Knop<sup>1,3\*</sup>

**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*<sup>1–5</sup>. 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 SWAp-Tag (SWAT) approach for high-throughput tagging of yeast ORFs and used it to N-terminally tag proteins of the endomembrane system<sup>6</sup>. 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<sup>6</sup> (Fig. 1a).

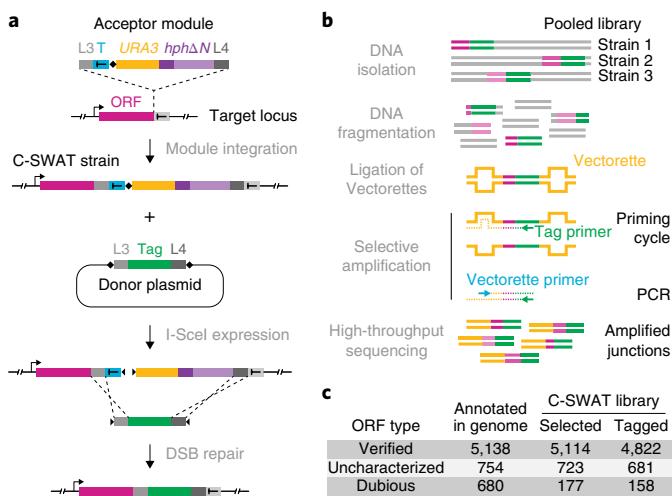
Here we developed a genome-wide C-SWAT (C-terminal SWAp-Tag) library. This arrayed library enables genome engineering at 3' ends of yeast ORFs and can be used for high-throughput C-terminal protein tagging. We used conventional PCR targeting<sup>7,8</sup> to insert a C-SWAT acceptor module before the stop codon of individual ORFs at endogenous chromosomal loci (Fig. 1a). The acceptor module consists of homology arms (L3 and L4, for subsequent recombination with the desired tag), a heterologous transcription terminator (T), a restriction site for the I-SceI endonuclease, a selection/counter-selection marker (*URA3*), and a second truncated selection marker (*hphΔN*) (Fig. 1a and Methods).

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<sup>9</sup> 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 sequence<sup>9</sup> (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<sup>6</sup> (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 ORFs<sup>10</sup>. 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. <sup>11</sup> and data

<sup>1</sup>Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH Alliance, Heidelberg, Germany. <sup>2</sup>Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel. <sup>3</sup>Cell Morphogenesis and Signal Transduction, DKFZ-ZMBH Alliance and German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>4</sup>Division of Redox Regulation, DKFZ-ZMBH Alliance and German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>5</sup>Present address: Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. <sup>6</sup>Present address: Department of Biophysical Chemistry, J. Heyrovsky Institute of Physical Chemistry of the CAS, Prague, Czech Republic. <sup>7</sup>Present address: Institute of Molecular Biology (IMB), Mainz, Germany. <sup>8</sup>These authors contributed equally: Matthias Meurer, Yuanqiang Duan, Ehud Sass. \*e-mail: a.khmelinskii@imb.de; emmanuel.levy@weizmann.ac.il; m.knop@zmbh.uni-heidelberg.de



**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 by 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 a functional protein). 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).

not shown). Thus, markerless tagging should preserve the gene-regulatory influence of sequences downstream of the tag.

The yeast GFP library<sup>1</sup>, comprising 4,159 ORFs tagged with GFP(S65T)<sup>12</sup>, has been widely used to study the yeast proteome. However, various fluorescent proteins with improved properties have been developed since its construction. The C-SWAT library provides a platform to profit from these developments. We found that in yeast, the green fluorescent protein mNeonGreen<sup>13</sup> and the red fluorescent protein mScarlet-I<sup>14</sup> were up to three times brighter than fluorescent proteins used in previous libraries<sup>15,6</sup> (Supplementary Fig. 3). Using the arrayed C-SWAT library, we tagged the yeast proteome with mNeonGreen and mScarlet-I, generating three new libraries: mNG-I (markerless tagging with mNeonGreen), mNG-II, and mSC-II (where mNeonGreen and mScarlet-I are followed by a heterologous terminator and the reconstituted *hph* marker) (Fig. 2a). We then quantified protein expression levels in these libraries using fluorescence measurements of colonies<sup>5</sup> (Methods). Over 4,300 proteins were expressed at detectable levels (>1.2-fold above background) in each library (Fig. 2b, Supplementary Fig. 4a, and Supplementary Table 2). This is consistent with the number of proteins detected by mass spectrometry in yeast grown under standard laboratory conditions<sup>15</sup>. Protein expression levels correlated well among the three libraries (Supplementary Fig. 4b) and with independent estimates of protein abundance<sup>15</sup> (Supplementary Fig. 4c), demonstrating the reliable nature of proteome-wide tagging with the C-SWAT library.

Next, we used the mNeonGreen libraries to examine how regulatory elements downstream of each ORF contribute to protein

expression. We observed that protein levels were on average ~20% higher in mNG-II strains, which carry the same heterologous terminator after the tag, than in mNG-I strains, in which the tag is followed by the endogenous terminator (Fig. 2c). Protein levels differed by more than twofold for ~11% of the proteome, with 466 and 10 proteins exhibiting lower and higher expression in mNG-I strains, respectively. Moreover, the difference between mNG-I and mNG-II strains correlated with the strength of the endogenous terminator for each ORF<sup>16</sup> (Fig. 2d). Consistent with these observations, the heterologous *ADH1* terminator used in mNG-II and mSC-II libraries is stronger than the terminator sequences of most yeast ORFs<sup>16</sup>. Taken together, these results demonstrate that tagging modules with heterologous terminators, commonly used for C-terminal protein tagging in yeast<sup>17,18</sup>, can measurably affect protein expression. It is likely that features present in the mRNA 3' UTR contribute to the observed effects, for example, by influencing mRNA stability or translation, which suggests that the C-SWAT library could be used to study the regulation of gene expression.

We observed expression of 207 proteins in the mNeon-Green libraries that were previously not detected with various approaches<sup>15</sup> (Fig. 2e). This group was enriched in ORFs that are annotated as uncharacterized (101 ORFs) or dubious (i.e., unlikely to encode functional proteins on the basis of available data; 93 ORFs)<sup>19</sup>. Using fluorescence microscopy, we detected expression of 72 out of 207 proteins even in mNG-I strains, where transcription of the tagged gene is not influenced by a heterologous terminator (Supplementary Table 3). Notably, 15 of these proteins (corresponding to 8 uncharacterized and 7 dubious ORFs) exhibited noncytosolic localization in both mNG-I and mNG-II strains (Fig. 2f,g, Supplementary Table 3, and Supplementary Data), which suggests that they are functional proteins. One example is the dubious ORF *YPR169W-A*. It is annotated as an unspliced variant of the uncharacterized ORF *YPR170W-B* (Supplementary Fig. 5a). mNeonGreen-tagged *Ypr169w-a* localized to the endoplasmic reticulum, whereas *Ypr170w-b* localized to the vacuolar membrane<sup>6</sup> (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*.

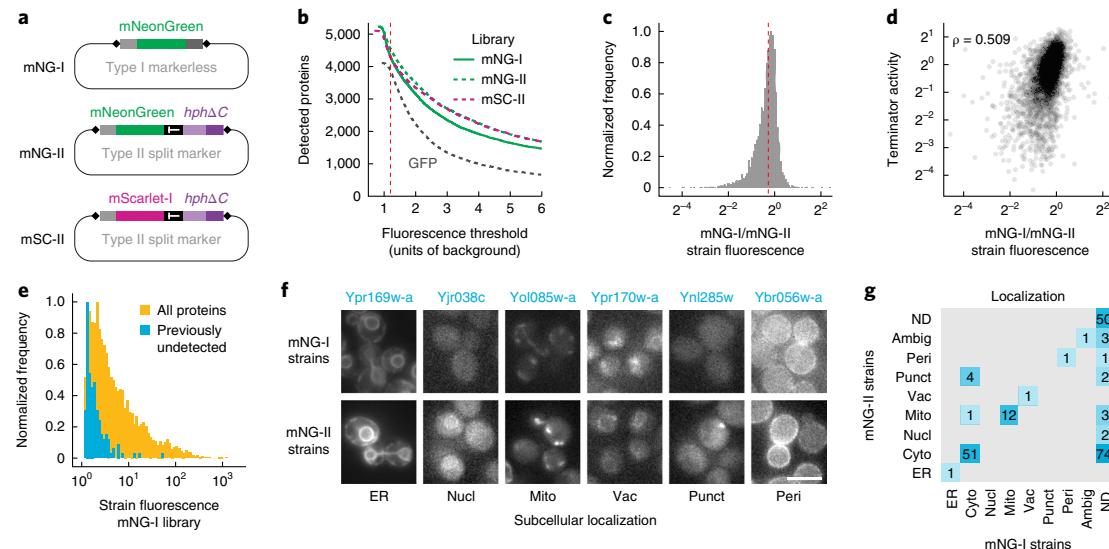
With updates to the yeast genome annotation, 80 ORFs included in the C-SWAT library have been reclassified from dubious to verified or uncharacterized since we began constructing the library (Supplementary Table 4). We note that the remaining dubious ORFs potentially also encode functional proteins, 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



**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<sup>20</sup> 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.<sup>16</sup>, and differences in protein expression levels between mNG-I and mNG-II strains. Spearman's rank correlation coefficient  $p = 0.509$ . **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

- Huh, W.-K. et al. *Nature* **425**, 686–691 (2003).
- Ghaemmaghami, S. et al. *Nature* **425**, 737–741 (2003).
- Gavin, A.-C. et al. *Nature* **440**, 631–636 (2006).
- Tarassov, K. et al. *Science* **320**, 1465–1470 (2008).
- Khmelinskii, A. et al. *Nature* **516**, 410–413 (2014).
- Yofe, I. et al. *Nat. Methods* **13**, 371–378 (2016).
- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. *Nucleic Acids Res.* **21**, 3329–3330 (1993).
- Wach, A., Brachet, A., Alberti-Segui, C., Rebischung, C. & Philippson, P. *Yeast* **13**, 1065–1075 (1997).
- Arnold, C. & Hodgson, I. J. *PCR Methods Appl.* **1**, 39–42 (1991).
- Khmelinskii, A., Meurer, M., Duishoeve, N., Delhomme, N. & Knop, M. *PLoS One* **6**, e23794 (2011).
- Bunina, D. et al. *Nucleic Acids Res.* **45**, 11144–11158 (2017).
- Tsien, R. Y. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Shaner, N. C. et al. *Nat. Methods* **10**, 407–409 (2013).
- Bindels, D. S. et al. *Nat. Methods* **14**, 53–56 (2017).
- Ho, B., Baryshnikova, A. & Brown, G. W. *Cell Syst.* **6**, 192–205 (2018).
- Yamanishi, M. et al. *ACS Synth. Biol.* **2**, 337–347 (2013).
- Longtine, M. S. et al. *Yeast* **14**, 953–961 (1998).
- Janke, C. et al. *Yeast* **21**, 947–962 (2004).
- Cherry, J. M. et al. *Nucleic Acids Res.* **40**, D700–D705 (2012).
- Chong, 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 (LDSF2). 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 HBGDS graduate school (I.K., K.H., and F.H.), an SFB1036 travel grant (E.S.), and the Alexander von Humboldt Foundation (M.Š.), 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., V.D., K.H., F.H., D.K., I.K., M.Š., 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](http://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'-cgtacgctgcaggtcgacgggtggcggttctggcggtggcgatcc-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'-agtctttttagatatcgattgaacgactcgaaattcatcgat-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 transcription<sup>21</sup>, which shows that this sequence is unlikely to interfere with regulatory elements downstream of ORFs.

**Library background strain.** The strain BY4741<sup>22</sup> (Supplementary Table 6), used to construct the most popular collections of yeast strains such as the knockout and GFP libraries<sup>23</sup>, 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 Y8205<sup>24</sup> (Supplementary Table 6). In this strain the *leu2A0* locus carries the *GAL1pr-NLS-I-SCE1-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 *can1A::STE2pr-SpHIS5* and *lyp1A::STE3pr-LEU2* markers for selection of *MATA* or *MATalpha* haploids at the end of the automated genetic crossing procedure via synthetic genetic array (SGA) methodology<sup>24</sup>.

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<sup>25</sup>. For donor type III, the *kanMX* selection marker was removed from the backbone. Tags can be cloned into these templates, for example, via BamHI + SpeI restriction enzyme cut sites. Here we constructed three donors for tagging with the mNeonGreen fluorescent protein (type I/II/III: pYD10/11/14) and one donor for tagging with the mScarlet-I fluorescent protein (pYD13, type II donor) (Supplementary Table 5). All plasmids and sequences are available from the Knop laboratory on request.

**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 (retrieved from the *Saccharomyces* Genome Database in July 2016). ORFs from the mitochondrial genome and the 2 $\mu$  plasmid were not included.

For each ORF, S2/S3 primers for PCR amplification of the tagging module were designed as previously described<sup>26</sup>. 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 study<sup>5</sup>. 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 5  $\mu$ M 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 (4titude, 4ti-0960) were filled with 40  $\mu$ l per well of a PCR mix using a Liquidator 96 channel manual pipettor (Mettler Toledo):

- 5  $\mu$ l of 10× HiFi-buffer (200 mM Tris-HCl, pH 8.8; 100 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 100 mM KCl; 1% (v/v) Triton X-100; 1 mg/ml BSA)
- 0.5  $\mu$ l of 100 mM stock of dNTPs (Bioline, BIO-39026)
- 0.15  $\mu$ l of 1 M stock MgCl<sub>2</sub>
- 5  $\mu$ l of 5 M stock of betaine (Sigma-Aldrich, 61962)
- 0.5  $\mu$ l of template DNA (200 ng/ $\mu$ l stock)
- 27.85  $\mu$ l of H<sub>2</sub>O
- 1  $\mu$ l of a high-fidelity DNA polymerase (1 U/ $\mu$ l)

A mixture of ORF-specific S2/S3 primers (10  $\mu$ l of 5  $\mu$ M stock) was added to each well from 96-well primer source plates with a 96-channel manual pipettor. The plates were sealed with aluminum seals (Steinbrenner Laborsysteme, SL-AM0550). PCR was then carried out in a Biometra TAdvance or TProfessional (Analytik Jena) under the following program:

- 2 min at 95 °C
- 30 cycles of 20 s at 95 °C/30 s at 66 °C/2.5 min at 72 °C
- 5 min at 72 °C
- incubation at 4 °C

Frozen BY4741 competent yeast cells were prepared and transformed with PCR-amplified acceptor modules in 96-well plates as previously described<sup>5,26</sup>. Transformation mixtures were manually plated onto 9-cm petri dishes with SC-Ura agar medium (synthetic complete medium lacking uracil and with 2% (w/v) glucose as a carbon source). After 2–3 d of incubation at 30 °C, six clones from each transformation were manually streaked for single colonies. Single colonies were then grown in 96-well format in SC-Ura medium with 15% (v/v) glycerol. The resulting six replicates of the C-SWAT library, one for each clone, were stored at -80 °C.

**Library validation by targeted sequencing (Anchor-Seq).** We developed and used Anchor-Seq to verify correct integration of the acceptor module in each C-SWAT strain. The validation procedure involved pooling of each replicate of the C-SWAT library, extraction of genomic DNA from each pool, DNA fragmentation and size selection, ligation of Vectorette adaptors, selective amplification of junctions between the C-SWAT acceptor module and upstream genomic sequences, and Illumina sequencing, as detailed below.

Each of the six C-SWAT library replicates was grown to saturation in 384-well plates (50  $\mu$ l of YPD medium per well). For each replicate, all strains were pooled, and cells were collected by centrifugation, washed once, resuspended in 10 ml of ddH<sub>2</sub>O, aliquoted, and stored at -80 °C until further processing.

Genomic DNA was extracted from each pool (300- $\mu$ l sample) with a Yeast Genomic DNA extraction kit (Zymo-Research, #D2002) according to the manufacturer's protocol with chloroform. The yield and quality of extracted DNA (typically 7–10  $\mu$ g) were assessed by absorbance at 260 and 280 nm wavelengths, measured with a NanoDrop ND-1000 spectrophotometer. Genomic DNA (5  $\mu$ g in 125  $\mu$ l of 10 mM Tris-HCl, pH 8.0) was fragmented in microTUBEs (Covaris, #520045) with a focused ultrasonicicator (Covaris, E220x) with shearing parameters set for a fragment size of 800 bp (shearing time of 50 s per tube, peak incident power of 105 W, duty factor of 5% and 200 cycles per burst).

Vectorette adaptors were then ligated to the sheared genomic DNA as follows. Single-stranded adaptor oligonucleotides (356-Vectorette and 355-Vectorette; Supplementary Table 7) were individually resuspended in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH, pH 7.5) to 100  $\mu$ M, mixed in equal amounts in a 1.5-ml Eppendorf tube and annealed (5-min incubation at 98 °C, cooling to ~23 °C in a water bath over a period of 3 h). Annealed Vectorette adaptors were then ligated to sheared genomic DNA (1  $\mu$ g) using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, #E7645S). End preparation and A-T ligation were performed according to the manufacturer's protocol with 15  $\mu$ M Vectorette adaptors. Fragment sizes of 300–800 bp were selected by agarose gel electrophoresis (3.5% NuSieve 3:1 agarose, Lonza #50090) and extracted with a Qiaquick gel extraction kit (Qiagen #28704) in 25  $\mu$ l of 10 mM Tris-HCl solution, pH 8.

DNA fragments containing the C-SWAT acceptor module were selectively amplified as follows. First, 20  $\mu$ l of purified adaptor-ligated fragments were used as input for 15 cycles of PCR with primers 503-Tag\_primer and 357-VC\_primer (Supplementary Table 7). The PCR product was purified with a Qiaquick PCR purification kit (Qiagen #28104) and used as input for 15 cycles of PCR with primers 382-P5 and 529-BC\_P7 (Supplementary Table 7). This set of primers added the P5 and P7 Illumina sequences and a six-nucleotide barcode for multiplexed sequencing. Fragment sizes were analyzed on a TapeStation 2200 (Agilent HS D1000 ScreenTape, #5067–5584) and typically followed a normal distribution peaking at 550–650 bp. In case fragments outside of 300–800 bp were present, samples underwent further size selection by agarose gel electrophoresis and gel extraction before Illumina sequencing. The outcome of selective amplification was controlled before sequencing, with quantitative PCR (qPCR) used to compare the abundance of fragments corresponding to ORFs that were tagged in the C-SWAT library (*ERG1*, *ERG11*, and *ERV25*; amplification with

an ORF-specific primer and a reverse primer in the acceptor module (Tag-rev); qPCR on-target (Supplementary Table 7) and ORFs that were not tagged (*ACO1*, *RPL30*, and *ASC1*; amplification with two ORF-specific primers; qPCR off-target; Supplementary Table 7). 5 ng of each sample (before and after selective amplification) were assayed using fast SYBR Green master mix (Thermo Fisher Scientific #4385612). 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 ( $\Delta CT$ ) and calculated as  $2^{(A+B)}$ , where  $A$  is the  $\Delta CT$  between off- and on-target sequences before selective amplification, and  $B$  is the  $\Delta CT$  for these sequences after selective amplification. We observed enrichments typically in the  $\sim 10^5$ – $10^6$ -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 (using primers 382-P5 and 529-BC\_P7) were used. To infer which genes were tagged correctly or incorrectly, we constructed expected reads using the 140 bp upstream of the stop codon for each ORF. Sequencing reads were then searched for matches to expected reads using custom Perl scripts. Given an ORF  $i$ , the search detected strict matches as well as six possible frameshifts ( $-3, -2, -1, +1, +2, +3$ ), and we denote the corresponding counts as  $M_{-3}$ ,  $F_{-3}^+$ ,  $F_{-2}^+$ ,  $F_{-1}^+$ ,  $F_{+1}^+$ ,  $F_{+2}^+$ , and  $F_{+3}^+$ . We also recorded cases where a strict match was found only for the first (1–70 bp upstream of the stop codon) or second half (71–140 bp upstream of the stop codon) of the sequence, presumably due to mutations that occurred in the half not matched. The corresponding counts are  $X_{-1}^1$  (first half) and  $X_{-1}^2$  (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  $M_{-1} > 5$ ,  $M_{-1} > (F_{-3}^+ + F_{-2}^+ + F_{-1}^+ + F_{+1}^+ + F_{+2}^+ + F_{+3}^+)$ ,  $5^*M_{-1} > X_{-1}^1$ , and  $5^*M_{-1} > X_{-1}^2$ .

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 arrayed in 96-well plates forming the C-SWAT library v1.0.

**Tag swap with the C-SWAT library.** Donor plasmids pMaM482, pMaM484, pYD10, pYD11, pYD13, and pYD14 (Supplementary Table 5) were transformed into the YM4639 donor strain.

To test the efficiency of tag swapping (Supplementary Fig. 2b), we selected 20 C-SWAT 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-SWAT library, which mostly contains C-SWAT 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<sup>15</sup> 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-SWAT 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-colony format according to the synthetic genetic array (SGA) procedure<sup>27</sup> as follows:

- mating of C-SWAT and donor strains on YPD plates (10 g/l yeast extract (BD Biosciences, 212750), 20 g/l peptone (BD Biosciences, 211677), 20 g/l glucose (Merck, 108337), 20 g/l agar (BD Biosciences, 214010)), 1 d at 30 °C
- selection of diploids on SC(MSG)-Ura+G-418 plates (1.7 g/l yeast nitrogen base without amino acids and ammonium sulfate (BD Biosciences, 233520), 1 g/l monosodium glutamic acid (MSG) (Sigma-Aldrich, G1626), 2 g/l amino acid mix SC(MSG)-Ura (glutamic acid replaced by MSG), G-418 (200 mg/l, Biochrom, A291-25), 20 g/l glucose, 20 g/l agar), 1 d at 30 °C
- sporulation on SPO plates (20 g/l potassium acetate (Sigma-Aldrich, 25059), 20 g/l agar), 5 d at 23 °C
- selection of haploids, step 1, on SC(MSG)-Ura/His/Arg/Lys+canavanine/thiavlysin plates (50 mg/l canavanine (Sigma-Aldrich, C1625), 50 mg/l thiavlysin (Sigma-Aldrich, A2636)), 2 d at 30 °C
- selection of haploids, step 2, on SC(MSG)-Ura/His/Arg/Lys+canavanine/thiavlysin/G-418 plates (50 mg/l canavanine, 50 mg/l thiavlysin, 200 mg/l G-418), 2 d at 30 °C
- selection of haploids, step 3, on SC(MSG)-Ura/His/Arg/Lys+canavanine/thiavlysin/G-418/clonNAT plates (50 mg/l canavanine, 50 mg/l thiavlysin, 200 mg/l G-418, 100 mg/l clonNAT (Werner BioAgents, 5.0)), 2 d at 30 °C

- induction of tag swapping on SC-His Gal/Raf plates (6.7 g/l yeast nitrogen base without amino acids (BD Biosciences, 291940), 2 g/l amino acid mix SC-His, 20 g/l galactose (Serva, 22020), 20 g/l raffinose (Sigma-Aldrich, R0250), 20 g/l agar), 2 d at 30 °C (done twice)
- selection against the acceptor module on SC-His+5-FOA plates (6.7 g/l yeast nitrogen base without amino acids, 2 g/l amino acid mix SC-His, 20 g/l glucose, 1 g/l 5-FOA (Apollo Scientific, PC4054)), 2 d at 30 °C

Finally, strains resulting from the swap of the full C-SWAT 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 SC(MSG)-His plates for all three donor types or to SC(MSG)-His + hygromycin plates (200 mg/l Hygromycin B Gold, Invivogen, ant-hg-5) for the type II donor or to SC(MSG)-His + G-418 plates (200 mg/l 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-SWAT library. Instead, manual pin tools can be used, as previously described<sup>27</sup>, 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  $\mu$  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 2–8  $\times 10^7$  cells/ml. Fluorescence measurements were performed on a BD FACSCanto RUO (BD Biosciences) equipped with a high-throughput sampler loader, a 488-nm laser and a combination of 505-nm long-pass and 530/30-nm band pass emission filters 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.

**Colony fluorescence measurements.** Strains resulting from the swap of the full C-SWAT 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 of 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. Scaled median absolute deviation (MAD) served as a robust estimate of s.d., and colonies within the 0.5th percentile of a normal distribution centered at the median 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<sup>28</sup>) and grown at 30 °C to mid-log phase for 7–8 h. 150  $\mu$ l of each culture were used for microscopy in glass-bottom 96-well plates (MGB096-1-2-LG-L; Matrical) coated with concanavalin A, as described<sup>29</sup>. Imaging was performed on a Nikon Ti-E widefield epifluorescence microscope with a 60× ApoTIRF oil-immersion objective (1.49-NA (numerical aperture); Nikon), an LED light engine (SpectraX, Lumencor), 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. Khmelinski, 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.

## ► Software

Policy information about [availability of computer code](#)

### 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.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 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

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

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

no animals used

Policy information about [studies involving human research participants](#)

### 12. Description of human research participants

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

no humans used