One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy

Ido Yofe[1,7], Uri Weill[1,7], Matthias Meurer[2], Silvia Chuartzman[1], Einat Zalckvar[1], Omer Goldman[1], Shifra Ben-Dor[3], Conny Schütze[4], Nils Wiedemann[4,5], Michael Knop[2,6], Anton Khmelinskii[2] & Maya Schuldiner[1]

The yeast *Saccharomyces cerevisiae* is ideal for systematic studies relying on collections of modified strains (libraries). Despite the significance of yeast libraries and the immense variety of available tags and regulatory elements, only a few such libraries exist, as their construction is extremely expensive and laborious. To overcome these limitations, we developed a SWAp-Tag (SWAT) method that enables one parental library to be modified easily and efficiently to give rise to an endless variety of libraries of choice. To showcase the versatility of the SWAT approach, we constructed and investigated a library of ~1,800 strains carrying SWAT-GFP modules at the amino termini of endomembrane proteins and then used it to create two new libraries (mCherry and seamless GFP). Our work demonstrates how the SWAT method allows fast and effortless creation of yeast libraries, opening the door to new ways of systematically studying cell biology.

Among the most important tools that the yeast model organism *S. cerevisiae* offers are systematic collections of strains (termed libraries) in which a different gene is modified in a similar manner in each strain to enable genome-wide studies[1]. Several libraries are currently available, such as a deletion library[2], a protein complementation assay library[3] and a GFP library[4]. Each established library has given rise to many biological insights[5–12]. But so far only very few libraries have been made, because the construction of each yeast library is an extremely laborious and expensive procedure. Researchers are deterred by the planning and ordering of thousands of primer sets, transformations, and clone-selection and validation steps.

We devised and implemented a methodology to remove the major hurdles of library construction for future library-assembly projects. Our methodology, termed SWAT, is based on an initial acceptor library that serves as a template that can be 'swapped' into other libraries of choice[13]. This requires the one-time construction of an acceptor library of strains (still relying on traditional methods[14–16]), each with an acceptor module integrated at a specific genomic location. The acceptor library can then be converted easily, rapidly and efficiently into any new library by replacement of the acceptor module with a new tag or genomic sequence of choice, introduced via crossing with a donor strain.

To establish this strategy, we constructed a library containing ~1,800 strains in which all proteins with known or predicted localization to the yeast endomembrane system were tagged with an amino-terminal (N′) SWAT acceptor module. The module used also contained a constitutive promoter and a GFP tag. With this N′-tag library, which we call the SWAT-GFP library, we were able to uncover the subcellular localization of hundreds of proteins never visualized before and to characterize several new peroxisomal and secreted proteins.

To demonstrate how easy it is to create a new library via the SWAT method, we created two additional libraries: a seamless GFP library in which the natural regulatory sequences were restored, and an mCherry library. These three N′-tag libraries are now freely available, along with a variety of donor and acceptor modules. We believe that the SWAT methodology will promote the construction of many new yeast libraries, opening the door to a flood of possibilities for systematic exploration of open questions in cell biology.

RESULTS

The SWAT strategy enables rapid library creation

The SWAT method relies on the genomic integration of an acceptor tagging module (using classical PCR-mediated transformation[14,15]) at a specific locus in a parental library. We made acceptor modules for tagging proteins at the N′ or C′ terminus, namely, N′-SWAT and C′-SWAT. The SWAT modules included a restriction site for the endonuclease I-SceI (that is not found elsewhere in the yeast genome)[17] and were flanked by two short (45–base pair (bp)) sequences that served as homology arms for recombination. These homology arms can also act as flexible protein linkers in cases where protein fusions are desired (L1, L2 for N′ tagging or L3, L4 for C′ tagging). Upon I-SceI induction, a genomic double-strand break is created within the SWAT module, facilitating its replacement

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. [2]Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH Alliance, Heidelberg, Germany. [3]Biological Services Department, Weizmann Institute of Science, Rehovot, Israel. [4]Institut für Biochemie und Molekularbiologie, ZBMZ Universität, Freiburg, Germany. [5]BIOSS Centre for Biological Signalling Studies, Universität Freiburg, Freiburg, Germany. [6]Cell Morphogenesis and Signal Transduction, Deutsches Krebsforschungszentrum (DKFZ), DKFZ-ZMBH Alliance, Heidelberg, Germany. [7]These authors contributed equally to this work. Correspondence should be addressed to I.Y. (ido.yofe@weizmann.ac.il), S.C. (silvia.chuartzman@weizmann.ac.il), A.K. (khmelins.zmbh@gmail.com) or M.S. (maya.schuldiner@weizmann.ac.il).

RECEIVED 22 MAY 2015; ACCEPTED 14 JANUARY 2016; PUBLISHED ONLINE 29 FEBRUARY 2016; DOI:10.1038/NMETH.3795

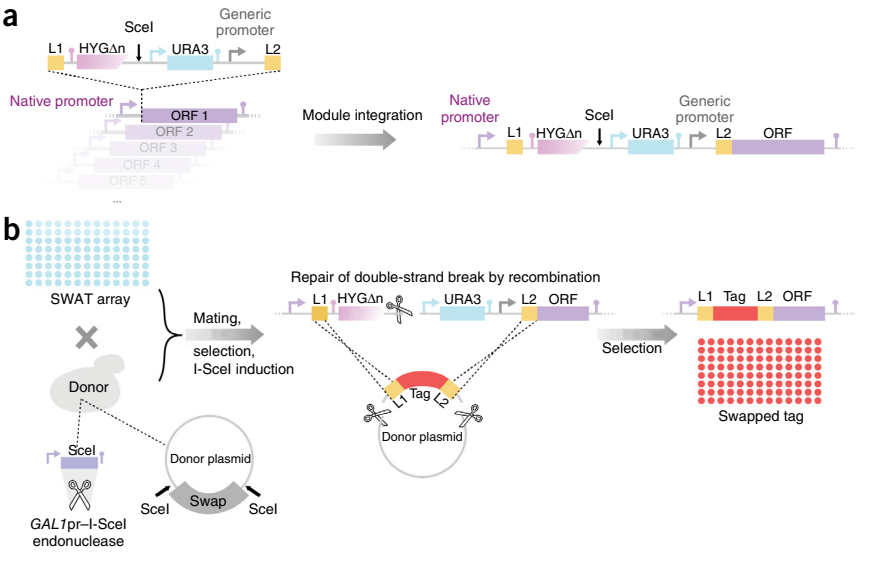

Figure 1 | The SWAT strategy enables fast and easy creation of systematic yeast libraries. (**a**) SWAT acceptor modules are used to tag proteins at the N terminus (**Supplementary Fig. 1** illustrates the approach for C′ tagging). The modules contain the restriction site for the I-SceI endonuclease (SceI), a *URA3* selection marker (URA3), a truncated hygromycin B selection marker (HYGΔn), and a generic promoter or terminator (depending on the terminus tagged), and they are flanked by two generic sequences for homologous recombination that also serve as linkers for protein fusion (L1, L2). (**b**) Left: an acceptor library of strains with SWAT-tagged genes (SWAT array) can be crossed with a donor strain to create a new library of choice. The donor strain encodes galactose-induced I-SceI (*GAL1*pr–I-SceI) and harbors a donor plasmid with the desired tag flanked by two generic sequences (L1, L2 or L3, L4) and two SceI restriction sites for systematic insertion. Middle: an automated mating procedure is used to create haploids carrying all genetic traits. At this point, I-SceI expression is induced by a growth step on galactose, leading to double-strand breaks in the tagged gene locus and the donor plasmid. The genomic breaks are mended by homologous recombination of the donor tag. Right: correct swapping is selected for, and a library featuring the new tag is created. (**c**) Several types of donor plasmid features (top) allow several options for creating tagged-gene libraries (bottom) (Online Methods).

by homologous recombination of any donor sequence that contains corresponding flanking regions. To select for such desired recombination events, we included two features in the SWAT tags: a *URA3* selection marker that allows both positive selection (on medium lacking uracil), for initial integration, and negative selection (on medium containing 5-fluoroorotic acid (5-FOA)), for module swap; and a truncated version of the hygromycin B resistance cassette (HYGΔ*n*) that can be restored to the full resistance gene by precise recombination (N′-SWAT in **Fig. 1a**, C′-SWAT in **Supplementary Fig. 1a** and both in **Supplementary Table 1**).

Once a library of strains is made in which the SWAT acceptor module is integrated into each gene, it is mated with a donor strain of the opposite mating type. The donor strain encodes for the I-SceI endonuclease under an inducible promoter (*GAL1*pr, which is induced by galactose and repressed by glucose) and harbors a donor plasmid that contains the desired sequence to be 'swapped in', also flanked by I-SceI sites. After automated synthetic genetic array[18] (SGA) procedures for mating, sporulation and selection of haploid spores of choice, I-SceI expression is induced by galactose-containing media, and the resulting recombination events are selected by one of various selection options (Online Methods). Starting with the acceptor library, the tag-swapping procedure is completed in approximately 3 weeks, after which a new library is formed and ready for experimentation (**Fig. 1b**).

A key advantage of the SWAT method is that any part of the acceptor module (promoter, tag or selection) can be replaced with any other; this includes replacement of the entire acceptor module to restore native gene regulation, allowing seamless tagging (N′-SWAT in **Fig. 1c**; C′-SWAT in **Supplementary Fig. 1b**, **Supplementary Fig. 2a**–**c** and Online Methods).

We measured the efficiency of swapping various modules and found that the SWAT method was highly efficient and accurate even for seamless swapping where no positive selection cassette was introduced (**Supplementary Fig. 2d**). To facilitate use by the community, we prepared an assortment of SWAT acceptor tagging modules and donor modules for either N′ or C′ tagging that can cater to a wide variety of needs; these are freely available (**Supplementary Tables 1** and **2**, **Supplementary Data 1** and http://www.weizmann.ac.il/molgen/Maya/SWAT).

Creation of N′ SWAT-GFP libraries

To demonstrate the SWAT approach, we constructed an acceptor library using an acceptor module that included the constitutive *SpNOP1* promoter[19] and a GFP tag (SWAT-GFP). We focused our efforts on endomembrane system proteins (residing on the surface or in the lumen of yeast secretory pathway organelles), which constitute nearly 30% of the yeast proteome[20].

We compiled a list of all endomembrane proteins and found that ~17% are predicted to contain an N′ cleavable signal peptide[21–24] (SP) that would be required for their correct targeting to the endoplasmic reticulum (ER) (**Supplementary Fig. 3**; a complete list of SP predictions is presented in **Supplementary Table 3**). To ensure that these proteins did not lose this essential targeting sequence, we created a unique SWAT acceptor module harboring the generic and efficient SP of Kar2 before the GFP tag (termed SWAT-SP-GFP) (**Fig. 2a**). The SP module was integrated downstream of the predicted cleavage site of each SP-containing protein (Online Methods).

We used the SWAT-SP-GFP module to tag 336 proteins with predicted SPs and succeeded with 318 of them (95%, SWAT-SP-GFP library). We used the SWAT-GFP module to tag the remaining 1,510 proteins that were not predicted to require a SP and succeeded in making 1,441 strains (95%, SWAT-GFP library). All together our libraries contain 1,759 strains that have also been

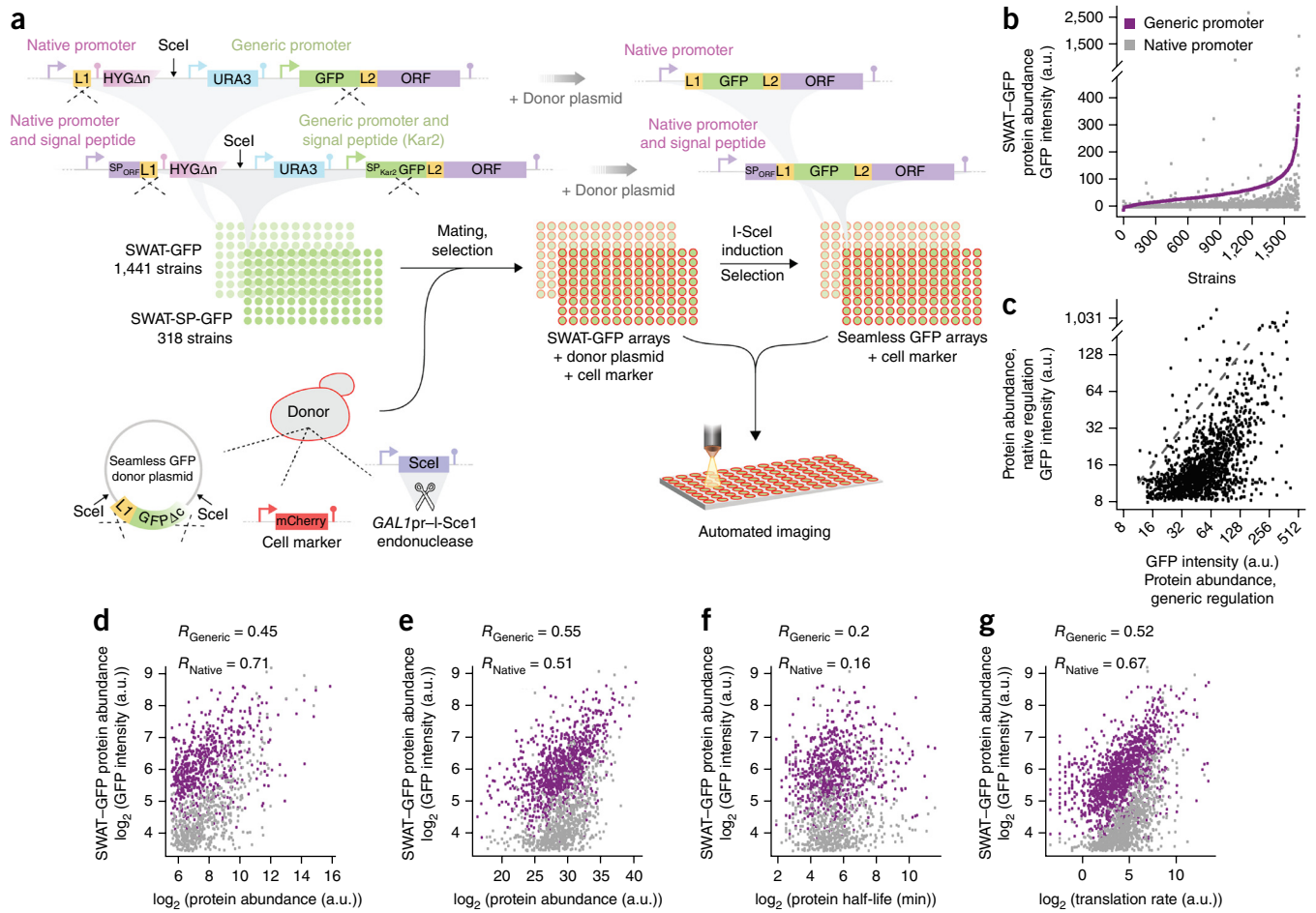

Figure 2 | Using the SWAT strategy to rapidly create a seamless N′-tagged GFP library enables comparison of protein abundance under generic or native regulation. (**a**) Workflow for library creation and imaging. The diagrams in the upper left represent two strain collections in which proteins have been tagged at the N′ terminus with an acceptor SWAT module containing the *SpNOP1* constitutive promotor (generic) and GFP. Proteins predicted to harbor a signal peptide (SP) were tagged with a similar acceptor SWAT module containing an SP upstream of GFP (SP_{Kar2}). The two libraries were crossed with a donor strain harboring a seamless GFP donor plasmid, the inducible I-SceI enzyme and an mCherry cell marker. After selection steps, the libraries were imaged before (SWAT-GFP arrays) and after (Seamless GFP arrays) tag swapping to compare protein abundance under generic versus native regulation. (**b**) Expression levels of GFP-protein fusions. The generic promoter represents SWAT-GFP, and the native promoter represents seamless GFP. (**c**) The correlation between protein abundance under generic and native regulation. The dashed line indicates the diagonal. (**d**–**g**) Comparison of N′ GFP–tagged protein abundances under either generic or native regulation showing the (**d**) abundance of these proteins as measured by flow cytometry analysis of C′-tagged GFP proteins[27], (**e**) abundance of these proteins as measured by mass spectrometry[28], (**f**) protein half-lives[29] and (**g**) protein translation rates as measured by ribosome profiling[30]. a.u., arbitrary units.

validated by PCR (**Fig. 2a** and **Supplementary Table 4**; further details on SP prediction, library construction and validation can be found in the Online Methods).

Once the SWAT-GFP libraries were assembled, we used them to perform high-throughput tag swapping with a 'seamless GFP donor' plasmid. The donor plasmid contained the L1 linker followed by 45 bp of GFP coding sequence starting from its first codon (GFPΔc) and was intended to excise all acceptor features, leaving only the short protein linker and the GFP tag. In the case of SWAT-GFP strains, this resulted in restoration of the native promoter, and in the case of the SWAT-SP-GFP strains it resulted in restoration of both the native promoter and the native SP (**Fig. 2a**). The swapping procedure was highly efficient: >98% of the strains were recovered, 96% of those tested (70/73) were accurately swapped (as demonstrated by PCR; **Supplementary Fig. 4**), and a homogeneous population was observed in >98% of strains tested (**Supplementary Fig. 5** and Online Methods). The end result of this procedure was an N′ seamless GFP library.

We used the GFP libraries before and after seamless swapping to measure how protein abundance is affected by regulatory sequences (native versus generic). Specifically, we measured GFP intensity at single-cell resolution using high-content microscopy[25,26]. Individual cells were identified by a cytosolic mCherry marker introduced by the donor strain (**Fig. 2a** and Online Methods). Surprisingly, we found that even when all proteins were under control of the *SpNOP1* constitutive promoter, each strain had a unique expression level spanning two orders of magnitude (**Fig. 2b**).

As expected, in the vast majority of cases the generic *SpNOP1* promoter produced higher expression than the native one. Whereas we could not detect a fluorescent signal above background autofluorescence for 2% of the proteins that were under regulation of the generic *SpNOP1* promoter, in the seamless

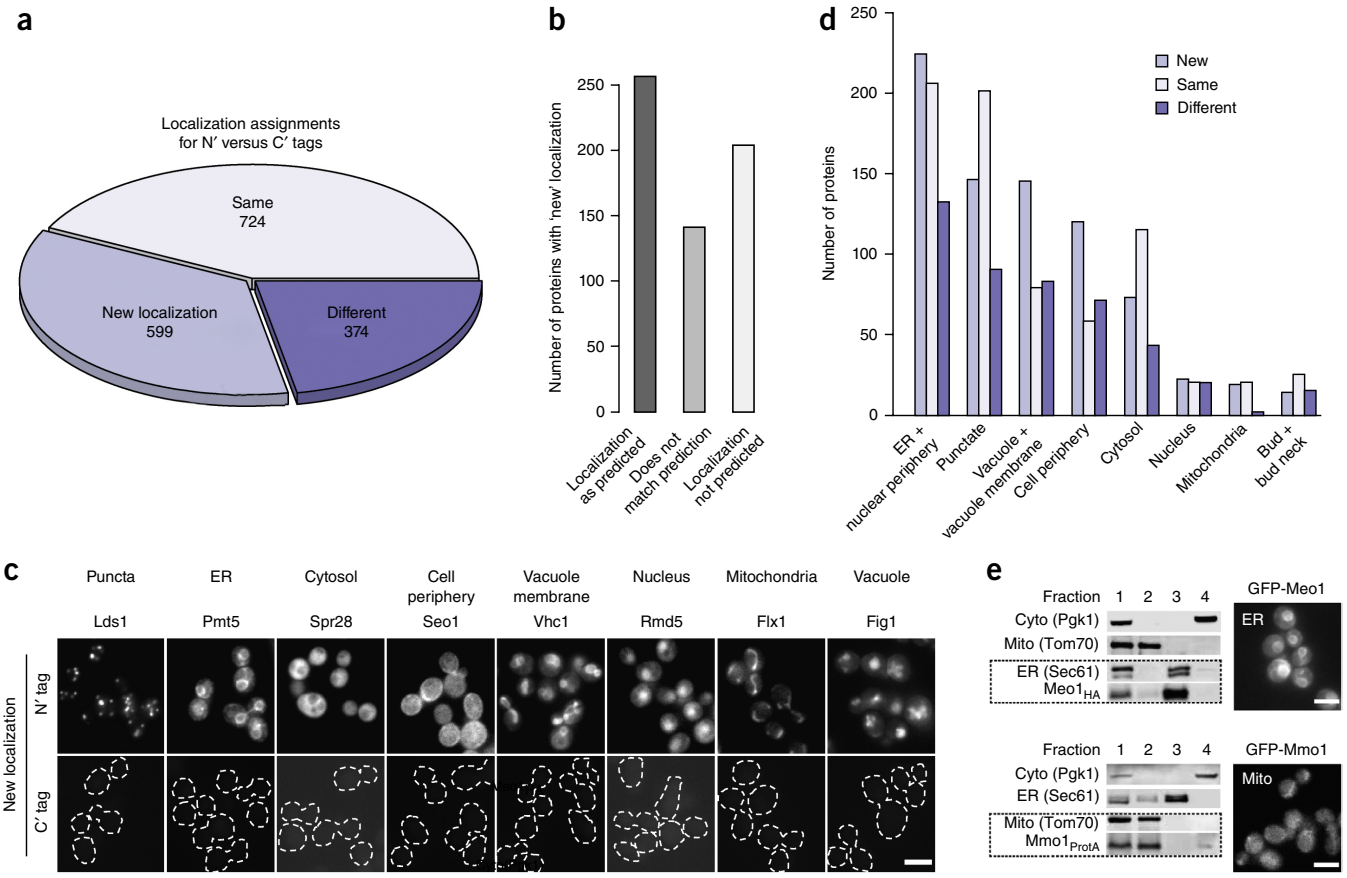

Figure 3 | Creation of an N′-SWAT library sheds new light on hundreds of endomembrane proteins. (**a**) Comparison of protein N′-tagging localization assignments to assignments made from the C′-tagged library[4]. A complete list of localization assignments is presented in **Supplementary Table 4**. (**b**) Comparison of new protein localization assignments to predicted localization by gene ontology. (**c**) Representative images showing proteins that reside in each subcellular localization from the 599 proteins (new localizations) that could be visualized by N′ tagging with the SWAT-GFP module (top) as opposed to C′ tagging with GFP (bottom). Scale bar, 5 µm (applies to all images in panel). (**d**) Protein-localization assignments by N′ tagging in the SWAT-GFP and SWAT-SP-GFP collections. (**e**) Western blots (left) and fluorescent images (right) of Ybr126w (newly named Meo1; top) and Ykl044w (newly named Mmo1; bottom). Western blots represent a fractionation assay. 1, postnuclear supernatant; 2, pellet (P13); 3, pellet (P100); 4, supernatant (S100) (Online Methods). Cyto, cytoplasm; mito, mitochondria. Scale bars, 5 µm.

version of the library this number increased dramatically to 48% (**Supplementary Table 4**). Although our analysis of swapping (above) suggests that ~4% of these strains simply did not undergo swapping, and although additional cases may represent inaccurate swapping, the majority represent cases in which the native promoter simply conferred low expression levels. This emphasizes the importance of having two versions of an N′ library—one that enables the study of proteins with naturally low abundance, and another that represents the native regulatory context.

The distinct correspondence between the abundance of proteins under the native versus the *SpNOP1* promoter highlights the importance of nontranscriptional events in regulating steady-state protein levels (**Fig. 2c**). When we compared protein abundance, under both generic and native regulation, to that in other, nonrelated data sets such as the C′ GFP library as measured by flow cytometry[27] (**Fig. 2d**) and mass spectrometry[28] (**Fig. 2e**), we noted high correlation. Whereas systematic data on protein half-lives[29] (**Fig. 2f**) did not seem to be highly correlated, data on protein translation rates by ribosome profiling[30] were (**Fig. 2g**), suggesting that translational regulation is the major source of abundance regulation for these proteins.

New condition-specific secreted proteins discovered

Our SWAT-SP-GFP library contained proteins with a predicted SP, and therefore it was possible that it included uncharacterized secreted proteins. To uncover such cases, we used a simple secretion blotting assay on both the acceptor SWAT-SP-GFP and the SP–seamless GFP libraries (**Supplementary Fig. 6**, **Supplementary Table 5** and Online Methods).

Clustering the strains in the SWAT-SP-GFP library by both secretion level and protein abundance highlighted several groups of proteins (**Supplementary Fig. 7a**). An interesting group of proteins seemed to be secreted only under the constitutive regulation of *SpNOP1pr* and the Kar2 SP. This could indicate that the overexpression of such proteins saturates retention and targeting mechanisms, thus causing aberrant secretion (as in the case of the vacuolar protein Prc1; **Supplementary Table 5**). However, another possibility is that some of the secreted proteins are *bona fide* secreted proteins that are usually not expressed under normal yeast growth conditions and are therefore secreted in a condition-specific fashion. Such condition-specific secreted proteins would not have been identified in the past, as mapping of secreted proteins has not been done under all possible yeast growth conditions.

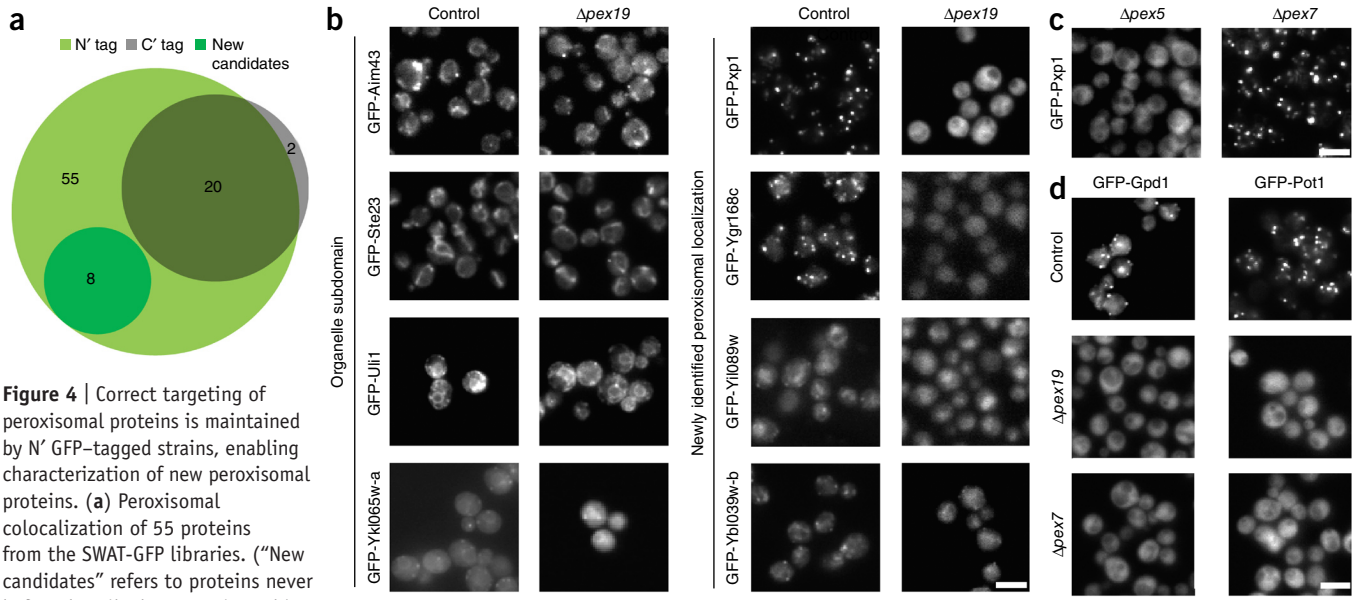

Figure 4 | Correct targeting of peroxisomal proteins is maintained by N′ GFP–tagged strains, enabling characterization of new peroxisomal proteins. (**a**) Peroxisomal colocalization of 55 proteins from the SWAT-GFP libraries. ("New candidates" refers to proteins never before described as peroxisomal.) (**b**) Fluorescent images of all eight newly identified peroxisomal candidates on a background of Δ*pex19*, which abolishes peroxisome biogenesis. Scale bar, 5 µm. (**c**) Fluorescent images of GFP-tagged Pxp1 on the background of a Δ*pex19* strain. Scale bar, 5 µm. (**d**) Fluorescent images of GFP-tagged Gpd1 and Pot1 on the background of Δ*pex19* and deletion of the PTS2 receptor, Δ*pex7*. Scale bar, 5 µm. In **b**–**d**, scale bars apply to all images in the respective panel.

We therefore focused on genes that were not expressed when under regulation of their native promoter but were very efficiently secreted once expression was driven by the generic promoter (**Supplementary Fig. 7b** and **Supplementary Table 5**). In this group, we saw that some of the proteins were indeed known condition-specific secreted proteins such as Pho5, which is secreted only during growth in low phosphate[31], and Bar1, which is secreted only from MATa cells[32]. Systematic analysis of Gene Ontology (GO)-term enrichment in the proteins of this cluster indeed showed overrepresentation of the "response to stimulus" category[33] ($P = 2 \times 10^{-4}$). Three proteins found in this cluster—Yil169c, Yfr020w and Yol159c—were previously uncharacterized, did not have a vacuolar localization (and so were not secreted simply because of saturation of retention machineries) and were not predicted to have transmembrane domains. We therefore named them Css1–3 (for condition-specific secretions 1–3) (**Supplementary Fig. 6b**). It is possible that these proteins have never been identified by traditional assays for finding secreted yeast proteins because the specific conditions required for their secretion remain to be uncovered.

Uncovering localization of hundreds of proteins

Reasoning that constitutive expression in the SWAT-GFP libraries might aid in the visualization of many proteins, we next annotated protein localization in all strains in the SWAT-GFP and SWAT-SP-GFP libraries, for both the acceptor and seamless versions (in at least two independently created clones to ensure reproducibility) (complete localization assignments are presented in **Supplementary Table 4**). Indeed, using the constitutively expressed SWAT-GFP library we were able to visualize 599 proteins for the first time (**Fig. 3a**). For nearly 200 of those proteins there was no prior information on localization in the literature, nor were there computational predictions, or manual or high-throughput annotations (**Fig. 3b**). Although the newly visualized proteins were found in all possible subcellular structures (**Fig. 3c**), we found that most new localization assignments were to the ER, punctate structures and vacuole membrane and lumen (**Fig. 3d**).

One group of proteins that had never been visualized systematically before is that of the small open reading frame proteins (smORFs) (fewer than 110 amino acids in length; a complete list of such smORFs is presented in **Supplementary Table 6**). This is because initial annotation of the yeast genome[34] did not include such ORFs, and they were annotated only later[35]. Our library contained 98 smORFs, of which only 39 had been visualized before. We were able to assign a localization to 78 of them (80%). To verify that the proteins had the capacity to be correctly localized with the GFP tag, we chose two examples of proteins that had not been studied previously: one of the 36 smORFs that localized to the ER (Ybr126w-a), and one of the 8 smORFs that localized to mitochondria (Ykl044w). We tagged the selected proteins at their C′ termini with a small epitope tag (hemagglutinin or ProtA) and used subcellular fractionation to verify their assigned localization. Indeed, Ybr126w-a was verified as a new ER protein that we named Meo1 (mini ER ORF1), and Ykl044w was verified as a new mitochondrial constituent that we named Mmo1 (mini mitochondria ORF1) (**Fig. 3e** and **Supplementary Fig. 8**). In the past, smORFs have been underrepresented in functional systematic libraries and therefore have been studied less than other yeast genes. We hope that the new information introduced here for both smORFs and other previously uncharacterized proteins (that now have a localization assignment), as well as the presence of these genes in our new libraries, will promote their study and the investigation of their function.

The N′ and C′ GFP libraries are complementary

Tagging proteins at either terminus can mask targeting sequences as well as regulatory sequences. We noted that 374 proteins were localized differently when tagged at their N′ terminus (in the SWAT-GFP libraries) versus their C′ terminus (as previously annotated[4,26]) (**Fig. 3a** and **Supplementary Fig. 9a**).

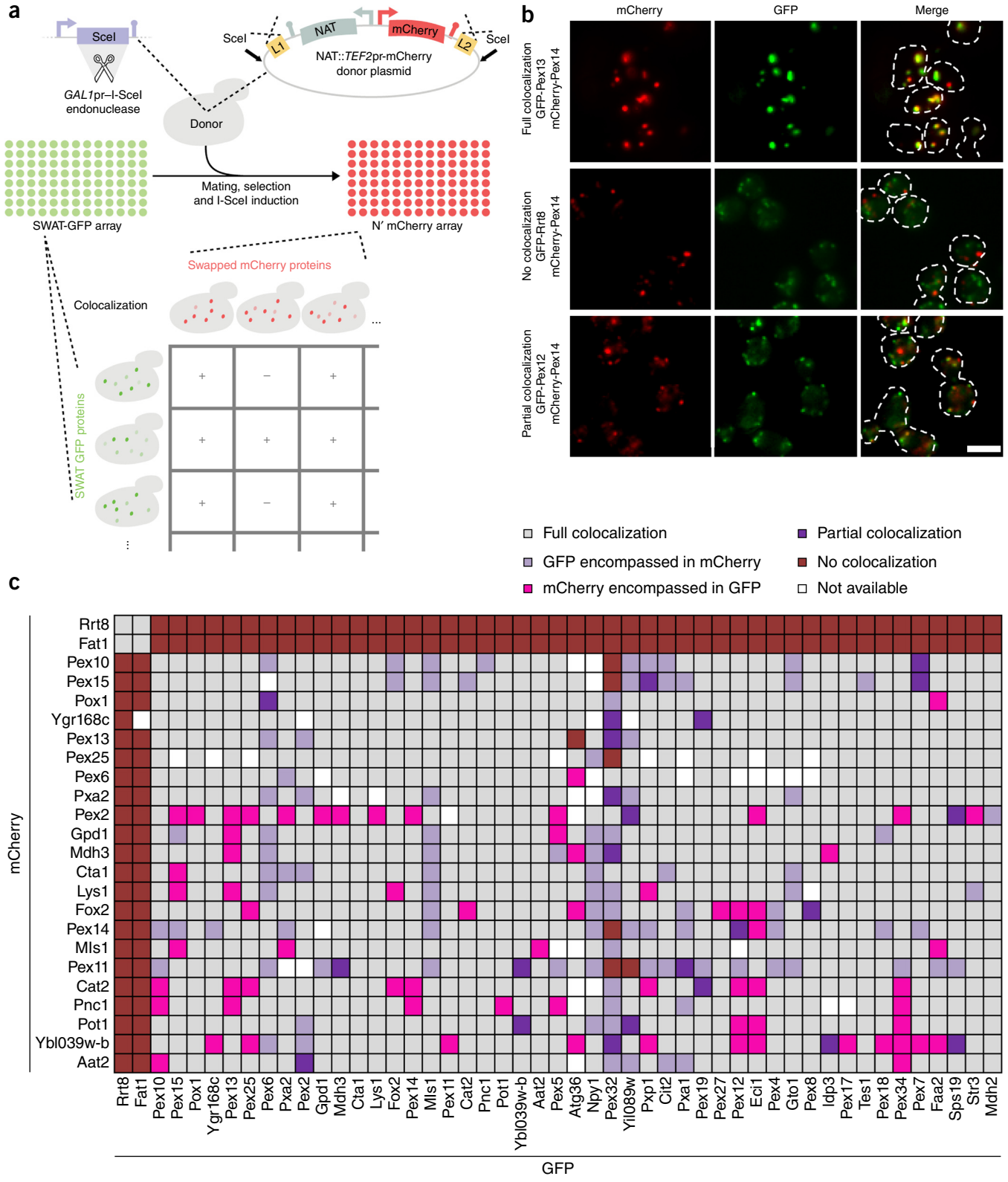

Figure 5 | Rapid creation of a new N′ mCherry library exemplifies the SWAT technology and opens up opportunities for systematic colocalization studies. (**a**) Workflow for creation of an N′ mCherry–tagged library. The SWAT-GFP arrays are crossed with a donor strain containing a donor plasmid with the swap cassette NAT::*TEF2*pr-mCherry (Online Methods). Strains from the mCherry library, selected as MAT-a, can then be crossed with MAT-a SWAT-GFP strains for systematic colocalization experiments. (**b**) Fluorescent images showing colocalization of the indicated strains. Scale bar, 5 µm (applies to all images in panel). (**c**) Peroxisomal protein colocalizations determined after mating of 24 mCherry-tagged strains with 49 GFP-tagged strains. The lipid droplet proteins Rrt8 and Fat1 served as negative controls.

Our endomembrane library encompasses many proteins that are known to be affected by masking of C′ localization signals. As expected, when we focused on such proteins (peroxisomal matrix proteins, tail- or GPI-anchored proteins, lipidated proteins and proteins with retrieval motifs) we found that many localized correctly only when tagged at their N′ terminus (**Supplementary Fig. 9b**). A comparison of the 374 differing localization assignments with manually curated GO annotations suggested that in about one third of the cases the localization seen with the N′ tag matches the GO annotation, in another third the C′ tag matches the GO annotation, and for the final third it is inconclusive (**Supplementary Fig. 10**). This demonstrates again how important it is to have a variety of complementary libraries to explore yeast cell biology.

Discovering new peroxisomal proteins

One organelle that was underrepresented in the C′ GFP library is the peroxisome. This is because many of the soluble proteins inside this organelle are targeted by virtue of the tripeptide targeting sequence PTS1 on the most C' of the protein[36], which becomes masked upon the addition of a C′ GFP tag. To uncover new peroxisomal proteins, we mated a strain expressing the peroxisomal marker Pex3-mCherry with the haploid SWAT-GFP and SWAT-SP-GFP libraries and assayed for colocalization in the emergent diploids in glucose. We found 55 proteins that colocalized with Pex3 (**Supplementary Table 7**). In contrast, only 22 proteins colocalized with Pex3 in the C′ GFP library[4], and we could identify all but 2 of them in the N′ library as well (**Fig. 4a**).

Eight proteins colocalized with Pex3 that had not previously been assigned as peroxisomal proteins (**Fig. 4a**). We recently showed that peroxisomes sit juxtaposed to the ER-mitochondria contact site[37]. Because the resolution attainable with light microscopy is not enough to discriminate between *bona fide* colocalization and juxtapositioning, we deleted Pex19, an essential protein for peroxisome biogenesis[38,39], from all eight GFP-tagged candidate peroxisomal proteins. Four proteins—Ykl065w-a, Uli1, Ste23 and Aim43—remained in a punctate structure even in the absence of mature peroxisomes (**Fig. 4b**) and therefore probably sit at peroxisomal contact sites on other organelles[40]. The other four proteins were redistributed to the cytosol upon Pex19 depletion, as expected from *bona fide* residents (**Fig. 4b**). These four new peroxisomal proteins had never been studied before and have only systematic names (Yel020c, Ygr168c, Yil089w and Ybl039w-b (which is also localized to another membrane)). One of the genes, *YEL020C*, has previously been suggested to be peroxisomal on the basis of computational predictions of PTS1 variants[41] and its presence in peroxisomal fractions during growth in oleic acid[42]. Yel020c is homologous to the main yeast pyruvate decarboxylase (Pdc1). We deleted the receptor for PTS1 targeting (Pex5) and, as a control, the receptor for PTS2 targeting (Pex7) and verified that Yel020c is targeted to peroxisomes by Pex5 (**Fig. 4c**). We therefore named this protein Pxp1 (peroxisomal protein 1).

An intriguing observation from our data was that the two proteins that are targeted to peroxisomes by the PTS2 system, Gpd1 and Pot1 (ref. 36), localized properly to peroxisomes (**Fig. 4d**). This was surprising because PTS2 targeting sequences are N-terminal and were thought to be dependent on position[43]. We deleted the PTS2 targeting receptor Pex7 and found that both proteins completely lost peroxisomal localization, proving that both proteins still used Pex7 to correctly target to peroxisomes and that the N′ tag does not interfere with the PTS2 signal. The ability to visualize the majority of peroxisomal proteins will enable future systematic efforts to study these organelles.

Rapid creation of a new N′ mCherry library

The most exciting feature of the SWAT technology is that the parental library can be rapidly swapped into any library with modifications of choice to address specific biological questions of interest. As an example, we decided to use the acceptor SWAT-GFP and SWAT-SP-GFP arrays to efficiently and easily switch out the green fluorophore and create a new library that contained mCherry-tagged proteins (**Fig. 5a** and Online Methods).

Creation of the N′ mCherry library again demonstrated the efficiency of the SWAT approach, as 98% of strains were retrieved after the procedure (**Supplementary Table 4**). We found that 83% of strains retained the same localization as in the SWAT-GFP and SWAT-SP-GFP parental libraries, which shows that the process was not only efficient but also accurate. We did, however, observe that the use of a stronger promoter (*TEF2pr*) in the donor module and the enhanced stability of mCherry in acidic environments created higher vacuolar background fluorescence[44].

We chose to use the N′ mCherry library to address an open question regarding peroxisome heterogeneity, namely, whether peroxisome subpopulations exist (**Fig. 5a**). To do this we selected 24 mCherry strains (22 peroxisomal proteins and two lipid droplet proteins serving as negative controls) and mated each of them with 49 GFP strains (47 peroxisomal proteins and the two lipid droplet proteins). We visualized the resulting 1,176 diploid strains and scored the extent of GFP-mCherry overlap (**Fig. 5a** and **Supplementary Table 8**). We found that peroxisomal populations were indeed heterogeneous (**Fig. 5b**,**c**). For example, many peroxisomal proteins (e.g., Pex10) appeared only in a subset of cellular peroxisomes (**Fig. 5c**). The mCherry library can now enable researchers to perform similar studies on other organelles that may have heterogeneous populations, such as mitochondria and lipid droplets.

DISCUSSION

The SWAT methodology opens a world of opportunities for studying cell biology through the creation of yeast libraries. The SWAT acceptor library can be manipulated to give rise to libraries harboring new regulatory sequences, RNA tags, protein processing signals, affinity-purification tags or protein fragment complementation tags. The ease and speed of this approach, coupled with its high efficiency and accuracy, should enable any yeast lab to create its own 'tailor-made' strain collection in a matter of a few weeks. Although robotic manipulation of libraries facilitates such efforts, all of these experiments can easily be undertaken using hand-operated pinning tools[25], removing technological barriers to SWAT library use. Another 4,500 strains must be made to complete the SWAT-GFP collection so that it is genome wide, and we are currently working toward that goal. However, this technology should be suitable for use in any other yeast species in which homologous recombination occurs at a high frequency. We hope that the SWAT method will advance the future of yeast research by boosting the number of systematic experimental possibilities.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

M.M. and S.C. contributed equally to this work. We thank N. Steinberg for his enormous contribution to the graphical design of this manuscript, and we thank E. Levy for exciting scientific discussions. This work was supported by the European Research Council (ERC) (Consolidator grants Peroxisystem 646604 to M.S. and MITOsmORFs 648235 to N.W.), Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centres (SFB 1140 to N.W. and SFB 1036, TP10, to M.K.) and the Excellence Initiative of the German Federal & State Governments (EXC 294 BIOSS to N.W.). We also acknowledge the generous support of the Mitzutani Foundation for Glycosciences (to M.S.), the Adelis Foundation (to M.S.) and the Kahn Center for Systems Biology (to M.S.).

AUTHOR CONTRIBUTIONS

A.K., M.M. and M.K. conceived the SWAT strategy. M.M., A.K., I.Y. and U.W. developed and implemented the method. S.B.-D. performed computational prediction of signal peptides. U.W., I.Y. and S.C. constructed the SWAT libraries (with help from O.G.) and performed all of the experiments and analysis on the systematic libraries. S.C. performed all automated robotic procedures. E.Z., C.S. and U.W. performed the low-throughput follow-up experiments. A.K., N.W., M.K. and M.S. supervised the work. I.Y., U.W., S.C. and M.S. wrote the manuscript with input from all other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Botstein, D. & Fink, G.R. Yeast: an experimental organism for 21st Century biology. *Genetics* **189**, 695–704 (2011).
2. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
3. Tarassov, K. *et al.* An in vivo map of the yeast protein interactome. *Science* **320**, 1465–1470 (2008).
4. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
5. Ben-Aroya, S. *et al.* Toward a comprehensive temperature-sensitive mutant repository of the essential genes of *Saccharomyces cerevisiae*. *Mol. Cell* **30**, 248–258 (2008).
6. Li, Z. *et al.* Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat. Biotechnol.* **29**, 361–367 (2011).
7. Mnaimneh, S. *et al.* Exploration of essential gene functions via titratable promoter alleles. *Cell* **118**, 31–44 (2004).
8. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
9. Breslow, D.K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5**, 711–718 (2008).
10. Kanemaki, M., Sanchez-Diaz, A., Gambus, A. & Labib, K. Functional proteomic identification of DNA replication proteins by induced proteolysis *in vivo*. *Nature* **423**, 720–724 (2003).
11. Sopko, R. *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
12. Sung, M.-K. *et al.* Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. *Genome Res.* **23**, 736–746 (2013).
13. Storici, F. & Resnick, M.A. The *delitto perfetto* approach to *in vivo* site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329–345 (2006).
14. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**, 3329–3330 (1993).
15. Wach, A., Brachat, A., Alberti-Segui, C., Rebischung, C. & Philippsen, P. Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae*. *Yeast* **13**, 1065–1075 (1997).
16. Yofe, I. & Schuldiner, M. Primers-4-Yeast: a comprehensive web tool for planning primers for *Saccharomyces cerevisiae*. *Yeast* **31**, 77–80 (2014).
17. Colleaux, L. *et al.* Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* **44**, 521–533 (1986).
18. Tong, A.H.Y. & Boone, C. High-throughput strain construction and systematic synthetic lethal screening in *Saccharomyces cerevisiae*. In *Yeast Gene Analysis* 2nd edn. (eds. Stansfield, I. & Stark, M.J.R.) 369–386, 706–707 (Elsevier, 2007).
19. Khmelinskii, A., Meurer, M., Duishoev, N., Delhomme, N. & Knop, M. Seamless gene tagging by endonuclease-driven homologous recombination. *PLoS One* **6**, e23794 (2011).
20. Fujita, M. & Kinoshita, T. GPI-anchor remodeling: potential functions of GPI-anchors in intracellular trafficking and membrane dynamics. *Biochim. Biophys. Acta* **1821**, 1050–1058 (2012).
21. Hegde, R.S. & Bernstein, H.D. The surprising complexity of signal sequences. *Trends Biochem. Sci.* **31**, 563–571 (2006).
22. Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
23. Käll, L., Krogh, A. & Sonnhammer, E.L.L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).
24. Reynolds, S.M., Käll, L., Riffle, M.E., Bilmes, J.A. & Noble, W.S. Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. *PLoS Comput. Biol.* **4**, e1000213 (2008).
25. Cohen, Y. & Schuldiner, M. Advanced methods for high-throughput microscopy screening of genetically modified yeast libraries. *Methods Mol. Biol.* **781**, 127–159 (2011).
26. Breker, M., Gymrek, M. & Schuldiner, M. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J. Cell Biol.* **200**, 839–850 (2013).
27. Newman, J.R.S. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
28. Picotti, P. *et al.* A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266–270 (2013).
29. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E.K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. USA* **103**, 13004–13009 (2006).
30. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
31. Barbarić, S., Münsterkötter, M., Svaren, J. & Hörz, W. The homeodomain protein Pho2 and the basic-helix-loop-helix protein Pho4 bind DNA cooperatively at the yeast PHO5 promoter. *Nucleic Acids Res.* **24**, 4479–4486 (1996).
32. Manney, T.R. Expression of the BAR1 gene in *Saccharomyces cerevisiae*: induction by the alpha mating pheromone of an activity associated with a secreted protein. *J. Bacteriol.* **155**, 291–301 (1983).
33. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
34. Basrai, M.A., Hieter, P. & Boeke, J.D. Small open reading frames: beautiful needles in the haystack. *Genome Res.* **7**, 768–771 (1997).
35. Kastenmayer, J.P. *et al.* Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* **16**, 365–373 (2006).
36. Hasan, S., Platta, H.W. & Erdmann, R. Import of proteins into the peroxisomal matrix. *Front. Physiol.* **4**, 261 (2013).
37. Cohen, Y. *et al.* Peroxisomes are juxtaposed to strategic sites on mitochondria. *Mol. Biosyst.* **10**, 1742–1748 (2014).
38. Götte, K. *et al.* Pex19p, a farnesylated protein essential for peroxisome biogenesis. *Mol. Cell. Biol.* **18**, 616–628 (1998).
39. Hettema, E.H., Girzalsky, W., van Den Berg, M., Erdmann, R. & Distel, B. *Saccharomyces cerevisiae* pex3p and pex19p are required for proper localization and stability of peroxisomal membrane proteins. *EMBO J.* **19**, 223–233 (2000).
40. Schuldiner, M. & Zalckvar, E. Peroxisystem: harnessing systems cell biology to study peroxisomes. *Biol. Cell* **107**, 89–97 (2015).
41. Schlüter, A., Real-Chicharro, A., Gabaldón, T., Sánchez-Jiménez, F. & Pujol, A. PeroxisomeDB 2.0: an integrative view of the global peroxisomal metabolome. *Nucleic Acids Res.* **38**, D800–D805 (2010).
42. Yi, E.C. *et al.* Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* **23**, 3205–3216 (2002).
43. Jung, S., Marelli, M., Rachubinski, R.A., Goodlett, D.R. & Aitchison, J.D. Dynamic changes in the subcellular distribution of Gpd1p in response to cell stress. *J. Biol. Chem.* **285**, 6739–6749 (2010).
44. Khmelinskii, A. & Knop, M. Analysis of protein dynamics with tandem fluorescent protein timers. *Methods Mol. Biol.* **1174**, 195–210 (2014).

ONLINE METHODS

Plasmid construction. We constructed plasmids using restriction-free cloning methods[45]. For a complete list of plasmids, see **Supplementary Tables 1** and **2** and **Supplementary Data 1**. The I-SceI restriction site sequence was agttacgctagg gataacagggtaatatag. The protein linker sequences (which also served as the generic recombination sites) were as follows: L1, 5′-cgtacgctgcaggtcgacggtggcggttctggcggtggcggatcc-3′; L2, 5′-ggcgg ttcctctggtggtggtggtgcgacagagaattcatcgatg-3′; L3, 5′-cgtacgctgca ggtcgacggtggcggttctggcggtggcggatcc-3′ (identical to L1); L4, 5′-ggcggttcctctggtggtggtggtgcgagctcgaattcatcgat-3′. Underlined sequences are the primer sequences used for amplification of the tagging module, corresponding to the pYM series sequences (S1, S4, S3 and S2, respectively)[46]. The use of these sequences ensured compatibility with existing oligo collections for these popular module sets.

The tagging modules included the constitutive promoter[19] *SpNOP1*pr to drive the fusion tag–protein expression. This promoter confers medium-level expression compared to stronger promoters such as *ScTEF1*pr and *ScGPD*pr.

The Kar2 SP sequence used in the SWAT-SP-GFP module was atgtttttcaacagactaagcgctggcaagctgctggtaccactctccgtggtcctgta cgccctttcgtggtaatattacctttacagaattctttccactcctccaatgttttagtta gaggtgccgat.

The codon-modified (to avoid altered recombination) Kar2 SP sequence used in the donor NAT::*TEF2*pr-SP_{Kar2}-mCherry plasmid was atgttcttcaatagattgtcagctgggaagcttcttgtgccactgtctgta gttctttacgcactgttcgtagtgatactacccctgcaaaactcctttcactcttcta atgtcctggtcagaggcgcagac.

Types of possible swapping options. *Seamless tag swap donor.* This type of donor contains the tag of choice flanked by the homologous recombination sequences (L1, L2 for N′ tag or L3, L4 for C′ tag). Use of this cassette eliminates all the basic acceptor components, giving rise to seamless fusion of the new tag (**Fig. 1c**; for an example, see **Supplementary Fig. 1a**). This swap is selected for using 5-FOA and leaves no positive selection marker.

Selection reconstitution and tag swap donor. This donor contains a new tag of choice flanked by one recombination sequence (L2 or L3, depending on whether N′ or C′ cassettes are used, respectively) and the N′ portion of the hygromycin B resistance cassette that both serves as homology for recombination and creates a new selection marker (**Fig. 1c**; for an example, see **Supplementary Fig. 1b**).

Selection and tag swap donor. This donor contains a new tag of choice and a new selection marker (not hygromycin B) flanked by homologous recombination sequences (L1, L2 for N′ tag or L3, L4 for C′ tag). Use of this cassette enables replacement of all constituents of the original SWAT module (selection cassette, regulatory sequences and tag) except for the linkers (**Fig. 1c**; for an example, see **Supplementary Fig. 1c**).

Signal peptide predictions of yeast proteins. We used three algorithms to predict the presence of a signal peptide in each yeast protein: SignalP 4.1 (ref. 22), Phobius 1.01 (ref. 23) and Philius[24]. We considered a protein as SP bearing for this study if all three programs predicted an SP in it, or if at least two programs predicted an SP with the exact same length (**Supplementary Table 3** and **Supplementary Fig. 2**).

Primer design. Primers for amplification of transformation cassettes and gene-specific targeting were designed with the Primers-4-Yeast web tool[16] (http://wws.weizmann.ac.il/Primers-4-Yeast) using the pYM plasmid type[46]. All tagging primers include a 40-bp homology sequence followed by 20 or 18 bp of cassette amplification sequence. The homology sequences were upstream and downstream of the protein start codon for normal N′ tagging and of the protein stop codon for C′ tagging, as described in the Primers-4-Yeast web tool. For N′ tagging of SP-containing proteins, homology sequences were designed to insert the cassette five amino acids downstream from the predicted SP cleavage point. Primers for validation of tagging and deletion transformations were also designed with the Primers-4-Yeast web tool, using the appropriate "Check primers" option. Primers were manufactured by Sigma-Aldrich in 96-well plates. A full list of primers used in this study is presented in **Supplementary Table 9**.

High-throughput yeast transformations. The BY4741 laboratory strain[47], which is the basis for most systematic yeast libraries, was used as the master strain for the collection. The SWAT-GFP and SWAT-SP-GFP acceptor modules (**Supplementary Table 1**; pST-N2 and pST-N3) were PCR amplified (KAPA Hi-Fi or KOD Hot Start DNA polymerase) in 96-well plates (Thermo Fisher Scientific) and transformed into BY4741. Transformations were carried out via a modified PEG-LiAc protocol[48] in a high-throughput manner. Each reaction was composed of 2.1 OD_{600} of cells (3 ml of cells at 0.7–0.8 OD_{600}), 120 µl of 50% PEG 3500 (wt/vol), 18 µl of 1 M LiAc, 25 µl of boiled SS-carrier DNA, 7 µl of double-distilled water and 20 µl of PCR-amplified transformation cassette DNA. Heat shock was applied in a PCR machine for 15 min at 30 °C followed by 30 min at 42 °C. Transformed cultures were plated on synthetic defined (SD)-URA media in 48-well divided agar plates (Bioassay X6029) and were incubated for 2–3 d at 30 °C. All procedures were carried out using an automated liquid handler (Janus, PerkinElmer).

Yeast strain validation and collection assembly. We chose at least four clones from each transformation. We performed validation PCR using a common forward primer from the 3′ end of the SWAT modules (S4 reverse complement) and a gene-specific reverse primer from the gene coding sequence. All four clones were imaged via a high-content screening platform in brightfield and GFP channels (see below). We reviewed images of all clones manually and assigned up to three localizations to each clone. Assignment categories were Ambiguous, Below Threshold, Bud, Bud Neck, Cell Periphery, Cytosol, ER, Mitochondria, Nuclear Periphery, Nucleolus, Nucleus, Punctate, Vacuole and Vacuole Membrane. Only strains with a duplicate repeating localization assignment that had been validated by PCR were chosen to compose the SWAT-GFP and SWAT-SP-GFP collections.

Analysis of swapping-procedure efficiency. We performed several tests to ascertain the consistency and efficiency of the SWAT strategy. First, we measured mCherry fluorescence levels by flow cytometry in SWAT-tagged strains (both C′ and N′ SWAT modules) that had undergone seamless mCherry swapping before and after I-SceI induction and after 5-FOA counterselection. In all cases mCherry introduction was demonstrated in >99% of cells (**Supplementary Fig. 1d**). Second, we performed PCR analysis

of the tagging state in several colonies of SWAT-GFP strains that had undergone seamless GFP swapping before and after I-SceI induction and after 5-FOA counterselection. The SWAT module was completely lost at the end of the process, resulting in seamless tagging (performed in a high-throughput manner) (an example is presented in **Supplementary Fig. 3**). Last, we measured the swapping efficiency of SWAT-GFP strains via single-cell GFP-expression analysis using microscopy. A homogeneous distribution of GFP expression was seen for >98% of strains after seamless GFP swapping (bimodal distribution test performed using the R dip test function; examples are visualized in **Supplementary Fig. 4**). It is worth noting that seamless swapping is expected to be impaired when the gene tagged with the SWAT module is located between identical or highly similar stretches of genomic DNA[19]. In such cases the double-strand break introduced by I-SceI in the SWAT module can be repaired by homologous recombination between the identical or similar genomic stretches, leading to excision of the entire genomic region instead of seamless swapping. Nevertheless, fewer than 250 of all yeast genes are expected to be affected by this, and therefore SWAT can be used for high-throughput seamless tagging of most of the proteome.

Donor strain construction. Donor strains were constructed on the background of an SGA[18] compatible query strain and contained a galactose-induced I-SceI endonuclease and a donor plasmid. To spare a selection marker in the donor strain, we introduced a *K. lactis URA3* selection marker into the *can1Δ* locus, upstream of the *STE2pr-SpHIS5* fragment (used for selection of MAT-a). A *Gal1pr-I-SceI* fragment was then introduced to replace the *URA3* selection, resulting in *can1Δ::GAL1pr-SceI::STE2pr-SpHIS5* (strain yMS2085). In cases where automated cell recognition software was required, we introduced a *NAT::TEF2pr-mCherry* cassette to the *hoΔ* locus. Examples of donor strain utilization can be found in **Figures 2a** and **5a**.

Automated manipulation of yeast libraries. We conducted automated strain maintenance and manipulation using a RoToR benchtop colony arrayer[25] (Singer Instruments). We carried out SGA procedures[18] for mating of the parental SWAT-GFP and SWAT-SP-GFP collections with donor strains bearing the seamless GFP donor (**Supplementary Table 2**; pSD-N9), the mCherry donor (**Supplementary Table 2**; pSD-N15/6) and the colocalization marker Pex3-RFP (**Fig. 4a**). After double mutant selection, tag swapping was prompted by two cycles of 1–2 d of growth on yeast extract peptone (YEP)-galactose (2%) media to induce I-SceI expression. Tag swapping was then selected by two cycles of growth on SD + 5-FOA (1 g/L) media for seamless GFP swap, yeast extract peptone dextrose (YEPD) + nourseothricin (NAT; 200 μg/mL) for the N′ mCherry swap, or YEPD + hygromycin B (200 μg/mL) for hygromycin B selection reconstitution swap.

Creation of mCherry library. To create the mCherry library, we built a donor plasmid of the 'selection and tag swap' type. The plasmid included the mCherry coding region under control of the *TEF2* promoter and a NAT selection marker flanked by the homology linkers and the I-SceI cut sites. For SP proteins we used a plasmid also containing an SP (**Fig. 5a**). We introduced the donor plasmids into the entire parental SWAT library and promoted tag swapping by induction of I-SceI and selection for 5-FOA– and NAT-resistant colonies. Using the SGA markers, we selected the resulting library of mCherry-tagged proteins such that all strains were of the opposite mating type from the original SWAT library (MAT-a).

High-throughput microscopy. We carried out high-content screening of strain collections using an automated microscopy setup (ScanR system, Olympus) as previously described[26]. We acquired images using a 60× air lens for GFP (excitation, 490/20 nm; emission, 535/50 nm), mCherry (excitation, 572/35 nm; emission, 632/60 nm) and brightfield channels. When a cytosolic mCherry cell marker was used (**Fig. 2**), images were analyzed using the ScanR Analysis software (Olympus), and single cells were recognized on the basis of the mCherry channel. Measures of cell size, shape and fluorescent signals were extracted. For localization assignments, we reviewed images manually using ImageJ. As we did not use any colocalization markers, we assigned only those localizations that could be easily discriminated by eye: ER, nuclear periphery, cytosol, cell periphery, vacuole lumen, vacuole membrane, mitochondria, nucleus, bud or bud neck, and punctate (which included the Golgi apparatus, peroxisomes, endosomes, other vesicular structures and subdomain compartments) (**Fig. 3c**).

Computational quantification of single-cell GFP intensity. We measured the median GFP intensity for each strain using single-cell recognition software (scanR Analysis software, Olympus) as previously described[26]. Strains with fewer than 30 recognized cells were excluded. We obtained the baseline autofluorescence level from strains not expressing GFP, and we classified any strain with median GFP expression below a cutoff of $\text{Mean}_{\text{autofluorescence}} + 2.58 \times \text{s.d. Mean}_{\text{autofluorescence}}$ as 'below threshold' (i.e., of detection).

Data processing. We compared the subcellular-localization annotations of the SWAT-GFP libraries (with the generic promoter and SP) with those of the C′-tag library (comprising data from two previously published data sets[4,26]). The pairwise comparisons between N′-tagging and C′-tagging annotations were classified in the following manner: 'Same' was assigned when at least one N′ annotation corresponded to a C′ one. 'New localization' was assigned if the C′ localization was classified as below threshold or ambiguous, or if no assignment existed. All other cases were classified as 'different' (**Fig. 3a**). We compared localization annotations of both N′- (SWAT-GFP) and C′-tagged proteins with GO cellular component manually curated annotations after applying the localization terminology to each GO term (**Fig. 3b** and **Supplementary Fig. 8**). All of the calculations were performed with either MATLAB or Rstudio. GO terms were retrieved from the *Saccharomyces* Genome Database (http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl).

Cluster analysis. We implemented agglomerative hierarchical clustering using MATLAB for the analysis of secretion levels (**Supplementary Fig. 7**) with a Euclidean or Cityblock distance metric for calculating the similarity between pairs of proteins.

Manual microscopy. We imaged specific strains in follow-up experiments with an Olympus IX71 microscope connected to a Photometrics CoolSNAP HQ camera, controlled by the Delta

doi:10.1038/nmeth.3795

Vision SoftWoRx 3.5.1 software, with 60× and 100× oil-immersion lenses using GFP (excitation, 490/20 nm; emission, 528/38 nm), mCherry (excitation, 555/28 nm; emission, 617/73 nm) and brightfield channels.

Protein-secretion assay. We analyzed protein secretion as previously described[49]. Specifically, yeast collections in either 384 or 1,536 colony array format were pinned by the RoToR colony arrayer onto YEPD plates, and a nitrocellulose membrane was applied onto the plate immediately thereafter. Overnight incubation at 30 °C allowed cell growth and ensured that secreted proteins were transferred to the nitrocellulose membrane. The membrane was removed from the plate and then washed (0.5 M NaCl, 10 mM Tris-HCl, pH 7.5) to release remaining cells. Secreted GFP-fused proteins were probed with a primary rabbit antibody to the GFP (custom-made by I. Braakman, Universiteit Utrecht, Utrecht, the Netherlands). Next we used secondary goat anti-rabbit conjugated to IRDye800 (926-32211, LI-COR Biosciences) and scanned membranes for infrared signal (including colony autofluorescence at 700 nm as a background control) using the Odyssey Imaging System (LI-COR Biosciences).

Subcellular fractionation. Yeast cells were grown to an OD_{600} of 1, and a total of 70 OD_{600} were harvested and washed in de-ionized H_2O. The pellet was suspended in 2 ml of DTT buffer (100 mM Tris-H_2SO_4, pH 9.4, 10 mM DTT) and incubated for 20 min at 30 °C. Spheroplasts were generated by incubation in 1 ml of Zymolyase buffer (1.2 M sorbitol, 20 mM potassium phosphate, pH 7.4, 0.5% Zymolyase (wt/vol)) for 45 min at 30 °C and harvested by centrifugation for 5 min at 3,000*g*. After being washed with 1 ml of 1.2 M sorbitol at 4 °C, spheroplasts were suspended in 2 ml of homogenization buffer (0.6 M sorbitol, 10 mM Tris-HCl, 1 mM EDTA, 1 mM PMSF, pH 7.4) and dounced 20 times in a cooled glass-Teflon potter on ice. The samples were subjected to a clearing spin for 5 min at 300*g* (4 °C), and half of the supernatant was used as postnuclear supernatant (PNS fraction 1), whereas the other half was further centrifuged for 15 min at 13,000*g* (4 °C) to generate a pellet (P13 fraction 2). The supernatant (S13) was centrifuged for 60 min at 100,000*g*, generating the P100 fraction 3 pellet and the S100 fraction 4 supernatant. Proteins in the PNS and S100 fractions were precipitated using StrataClean resin (Agilent Technologies).

SDS-PAGE, western blotting and immunodecoration. After Tricine–SDS-PAGE, proteins were transferred to PVDF (polyvinylidene difluoride) membranes using the semidry-transfer method[50]. Proteins blotted onto PVDF membranes were detected by immunodecoration with custom-made specific primary antibodies (Tom70, GR657-5; Pgk1, GR753-7; Sec61, GR760-1) and horseradish peroxidase–coupled secondary antibodies (hemagglutinin, 12013819001, Roche; peroxidase, P1291, Sigma-Aldrich; rabbit immunoglobulin-γ, 111-035-003, Jackson ImmunoResearch) after incubation with chemiluminescence solution and recorded using the LAS-4000 CCD (charge-coupled device) camera system and MultiGauge (Fujifilm).

Obtaining the libraries, plasmids and protocols. All strains, plasmids and libraries presented in this manuscript are freely available upon request. All protocols for using the SWAT strategy can be downloaded from our lab website (http://www.weizmann.ac.il/molgen/Maya/SWAT).

45. Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R. & Peleg, Y. Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.* **172**, 34–44 (2010).
46. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
47. Brachmann, C.B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
48. Gietz, R.D. & Woods, R.A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).
49. Copic, A. *et al.* Genomewide analysis reveals novel pathways affecting endoplasmic reticulum homeostasis, protein modification and quality control. *Genetics* **182**, 757–769 (2009).
50. Schägger, H. Tricine–SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).