nature methods

ARTICLES

https://doi.org/10.1038/s41592-018-0044-9

Genome-wide SWAp-Tag yeast libraries for proteome exploration

Uri Weill[1,10], Ido Yofe[1,10], Ehud Sass[2], Bram Stynen[3], Dan Davidi[4], Janani Natarajan[5], Reut Ben-Menachem[6], Zohar Avihou[1], Omer Goldman[1], Nofar Harpaz[1], Silvia Chuartzman[1], Kiril Kniazev[1], Barbara Knoblach[7], Janina Laborenz[8], Felix Boos[8], Jacqueline Kowarzyk[3], Shifra Ben-Dor[9], Einat Zalckvar[1], Johannes M. Herrmann[8], Richard A. Rachubinski[7], Ophry Pines[6], Doron Rapaport[5], Stephen W. Michnick[3], Emmanuel D. Levy[2] and Maya Schuldiner[1]*

Yeast libraries revolutionized the systematic study of cell biology. To extensively increase the number of such libraries, we used our previously devised SWAp-Tag (SWAT) approach to construct a genome-wide library of ~5,500 strains carrying the SWAT *NOP1promoter-GFP* module at the N terminus of proteins. In addition, we created six diverse libraries that restored the native regulation, created an overexpression library with a Cherry tag, or enabled protein complementation assays from two fragments of an enzyme or fluorophore. We developed methods utilizing these SWAT collections to systematically characterize the yeast proteome for protein abundance, localization, topology, and interactions.

Among the most important tools that the model organism *Saccharomyces cerevisiae* (yeast) offers are systematic collections of strains, or libraries, in which each gene is modified in a similar manner to enable genome-wide studies[1–4]. We recently developed a methodology termed SWAT[5,6] for the rapid creation of tagged yeast libraries. SWAT acceptor libraries enable the replacement of the acceptor module with a new tag or genomic sequence of choice, introduced via crossing with a donor strain. The resulting libraries can then be used for systematic assays or as a strain reservoir for individual protein studies[5].

We created, to the best of our knowledge, the first whole-genome SWAT library with an N′ tag (for the complementary C′ SWAT library, see ref. [7]), covering ~90% of yeast genes[8]. Imaging of this library allowed us to determine the localization of 796 yeast proteins that could not be visualized before with a C′ fluorescent tag. We constructed six additional libraries to explore many aspects of yeast cell biology: the role of promoters in regulation of protein expression, the mitochondrial protein roster, protein interactions on a whole-organelle level, and systematic assessment of protein topology.

Results

Generation of a SWAT full-genome collection. We compiled the sequences of all yeast genes using the *S. cerevisiae* genome database (SGD)[8] and attempted to tag 3,916 proteins to complete the N′, genome-wide SWAT library with the *NOP1pr-GFP* tag (which encodes Superfolder GFP[9]) (Supplementary Table 1). Because the SWAT cassette is added to the N′ of proteins, we previously used a tailored cassette encoding a strong signal peptide (SP) for all endomembrane system proteins that harbor such a targeting signal[5]. In the current extended library, we added a cassette encoding a strong mitochondrial targeting signal (MTS) to the several hundred proteins that have such a targeting signal at their N′ (*NOP1pr-MTS-GFP*) (Fig. 1a and Supplementary Table 2). Overall, the final SWAT full-genome library contains 5,457 strains that underwent several quality control steps (Supplementary Tables 1 and 3 and Methods).

Using the SWAT approach to analyze the role of promoters in regulating protein abundance. We first used our library to investigate the relative contribution of promoters to protein expression levels. For this, we created two additional full genome libraries: a native promoter GFP library termed NATIVEpr-GFP, which restores the natural promoter and endogenous 5′ untranslated regions (UTRs) (as well as the native MTS or SP for the relevant proteins; Fig. 1a and Supplementary Fig. 1), and a TEF2pr-mCherry library that introduces one of the strongest promoters in yeast[10,11]. In addition, the TEF2pr-mCherry library provides a collection that is tagged with a different fluorophore, thus enabling colocalization studies[5].

We imaged all three libraries with an automated microscopy system and analyzed each strain for fluorescence intensity. To our surprise, in both the TEF2pr-mCherry and NOP1pr-GFP libraries, proteins had extremely diverse expression levels, spanning over two orders of magnitude (Supplementary Fig. 2a), despite harboring an identical promoter. This highlighted that the promoter was responsible for only a fraction of the regulation involved in expression.

Indeed, when we compared expression levels of proteins in the NATIVEpr-GFP library, we found they had no more of a correlation to strains that preserve the native promoter, such as the

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. [2]Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel. [3]Département de Biochimie, Faculté de Médecine, Université de Montréal, Montreal, QC, Canada. [4]Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel. [5]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany. [6]Department of Microbiology and Molecular Genetics, IMRIC, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel. [7]Department of Cell Biology, University of Alberta, Edmonton, AB, Canada. [8]Department of Cell Biology, University of Kaiserslautern, Kaiserslautern, Germany. [9]Department of Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel. [10]These authors contributed equally: Uri Weill, Ido Yofe.
*e-mail: maya.schuldiner@weizmann.ac.il

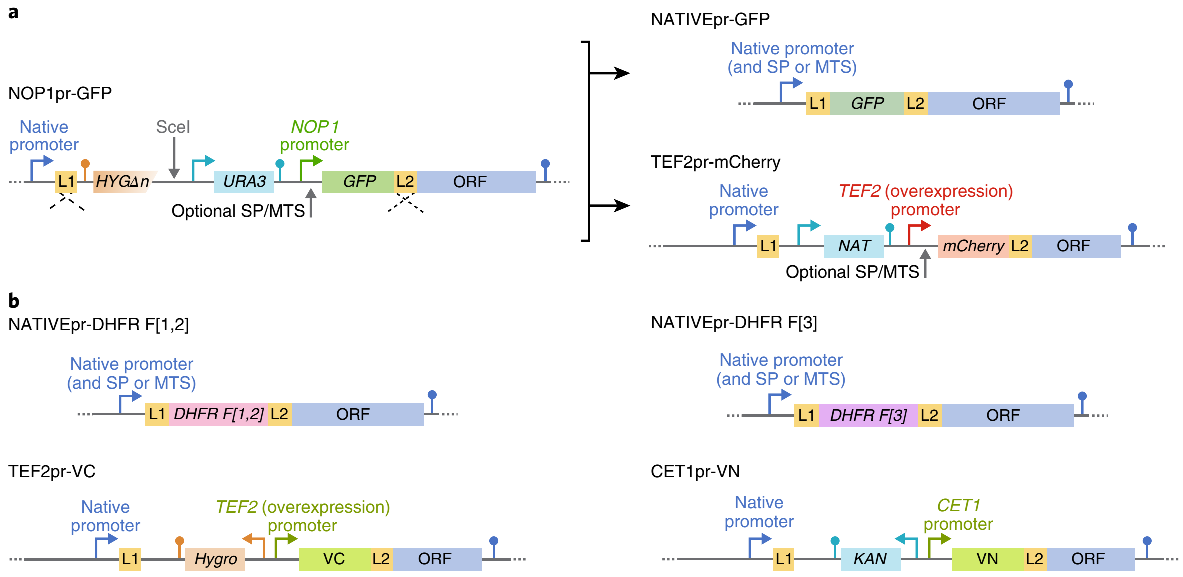


Fig. 1 | Library generation of genome-wide N′ yeast collections by SWAT technology. a, Composition of the N-terminal SWAT cassette harboring the *NOP1* constitutive promotor and GFP. For the relevant subset, the tagging cassette also encoded an MTS upstream of GFP (MTS$_{Su9}$) or an SP[5] upstream of GFP (SP$_{Kar2}$). The SWAT parental library underwent swapping to either a NATIVEpr-GFP (with the native MTS or SP restored as well) or TEF2pr-mCherry. **b**, The SWAT parental library was also swapped to create four additional libraries: TEF2pr-VC, CET1pr-VN, NATIVEpr-DHFR F[1,2], and NATIVEpr-DHFR F[3].

C′ GFP library[12] (correlation of 0.43; Supplementary Fig. 2b), or native abundance as measured by mass spectrometry[13] (correlation of 0.56; Fig. 2b), relative to the abundance of proteins under the *NOP1* promoter (correlation of 0.58; Fig. 2a) or the *TEF2* promoter (correlation of 0.42; Fig. 2b).

To identify what other elements affect protein abundance, we compared our observed protein abundance measurements to published systematic datasets, such as mRNA abundance as measured by RNA-seq[14], protein translation rates as measured by ribosome profiling[15], mRNA half-lives[16], and protein half-lives[17] (Fig. 2b). Little correlation was found to RNA or protein half-lives. Poor correlation may be a result of different experimental setups, but can also suggest that RNA and protein degradation may act as a regulatory mechanism affecting the abundance of specific proteins, rather than acting globally. We observed the highest correlation with mRNA abundance and translation rates, which suggests that chromatin state[18], together with translation, acts as a global effector of abundance regulation across the proteome.

The SWAT-GFP library identifies a cellular localization for hundreds of proteins. We then annotated protein localization in all strains of the NATIVEpr-GFP, NOP1pr-GFP, and TEF2-mCherry libraries. Assignments were given only to organelles or cellular locations that could be unequivocally determined without the need for colocalization (Supplementary Table 1). Given that punctate localization can represent a variety of compartments that can be distinguished only by colocalization studies[3,19], we dubbed such proteins ‘punctate’.

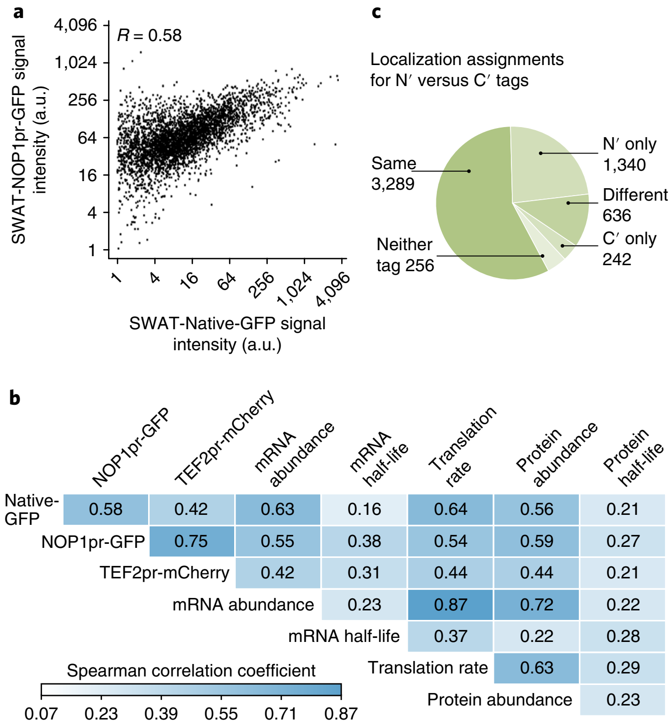
We then compared the current tally of localizations in the N′ NOP1pr-GFP library to previously assigned localizations with the C′ GFP library[3,12] (Fig. 2c). We found that 3,289 proteins showed the same localization, strongly supporting the previously assigned location as the correct one for these proteins. Other proteins (242) could be localized only in the C′ library or could not be tagged or visualized in either N′ or C′ libraries (256). Of these, 63 are essential for viability[2]. Many proteins (636) displayed a localization that was different between N′ and C′ tagging. Additional work will be required to distinguish whether one tag is superior or whether both locations or neither location is correct. Notably, we were able to assign a localization for an additional 796 proteins that have not been previously visualized in libraries. Taking into consideration the 544 new localizations from our previous study[5], a total of 1,340 protein localizations were assigned on the basis of the NOP1pr-GFP library.

Using the SWAT libraries to define a more complete mitochondrial proteome. Mitochondria in yeast have already been assigned over 900 high-confidence resident proteins[20]. Our N′ libraries provided an opportunity to complete the mitochondrial proteome roster through visualization of mitochondrial proteins that were not recognized before because of either tag interference or conditional expression.

To ensure that we would capture the maximal repertoire of mitochondrial proteins, we built a comprehensive list of proteins with either a high-probability predicted MTS (according to Mitofates[21] and TargetP version 1.1[22]) or an experimentally verified one[23,24] (Supplementary Table 2), and created an N′ SWAT library covering 359 of the 420 proteins that were designated by our analysis as having an MTS (a complete list is presented in Supplementary Table 1). MTS-containing proteins were tagged using a specific N′-tagging cassette that has a generic *Su9*-MTS before the GFP tag inserted downstream of the MTS cleavage site (similarly to the SP-containing proteins in the endomembrane SWAT library[5]). With the cassette inserted 15 nt (five amino acids) downstream of the original MTS cleavage site, the synthetic MTS was able to direct the protein into mitochondria and, after being cleaved, would leave the GFP moiety fused to the mature protein.

We first verified that the *Su9*-MTS was sufficient to establish a mitochondrial localization (Supplementary Fig. 3a) and that the *NOP1pr-MTS-GFP* cassette supported the SWAT approach to return to a native MTS and promoter (Supplementary Fig. 3b).

We previously found that N′ tagging of predicted mitochondrial proteins without an MTS can reveal mitochondrial localization for even very low-abundance proteins[20]. An additional 15 new mitochondrial proteins that did not have an MTS were



	NOP1pr-GFP	TEF2pr-mCherry	mRNA abundance	mRNA half-life	Translation rate	Protein abundance	Protein half-life
Native-GFP	0.58	0.42	0.63	0.16	0.64	0.56	0.21
NOP1pr-GFP		0.75	0.55	0.38	0.54	0.59	0.27
TEF2pr-mCherry			0.42	0.31	0.44	0.44	0.21
mRNA abundance				0.23	0.87	0.72	0.22
mRNA half-life					0.37	0.22	0.28
Translation rate						0.63	0.29
Protein abundance							0.23

Fig. 2 | Genome-wide N′ tagged collections enable investigation of abundance regulation and localization assignment. a, Scatter plot showing the correlation between protein abundance of NATIVEpr-GFP-tagged strains and protein abundance under generic regulation (NOP1pr-GFP). *R* represents two-sided Spearman correlation test score. a.u., arbitrary units. **b**, Spearman correlation scores of the expression levels of fluorophore–protein fusions of the TEF2pr-mCherry, NATIVEpr-GFP, and NOP1pr-GFP libraries with respect to mRNA abundance as measured by RNA-seq[14], protein translation rates as measured by ribosome profiling[15], mRNA half-lives[16], protein half-lives[17], and protein abundance as measured by mass spectrometry[13]. **c**, Comparison of N′ NOP1pr-GFP library localization assignments to assignments made from the C′-tagged library[3,12]. A complete list of localization assignments is presented in Supplementary Table 1. Quantitation of abundance was preformed once. Strains with a final abundance score less than 1 were excluded from the data analysis and Spearman correlation tests.

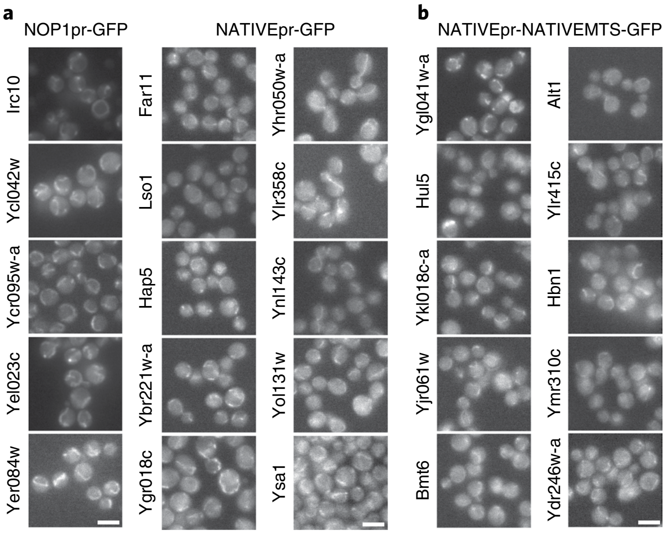


Fig. 3 | Characterization of the mitochondrial proteome. a, Visualization of mitochondrial proteins not containing an N′ MTS with N′-tagged NOP1pr-GFP (left) or NATIVEpr-GFP (right). **b**, Visualization of mitochondrial proteins seen with the NATIVEpr-(NATIVE)MTS-GFP library (not including Su9-MTS). Scale bars, 5 μm. Imaging of strains was performed a single time. Images represent the entire field.

found in the whole-genome library either when tagged with the *NOP1pr-GFP* cassette (not including Su9-MTS) or in the native promoter version (Fig. 3a). We verified one such new protein, Ysa1 (Supplementary Fig. 4).

We also verified mitochondrial localization for several of the MTS-containing proteins. Because the *Su9*-MTS is dominant and could mistarget nonmitochondrial proteins into mitochondria, we assigned mitochondrial localization to proteins only if we could verify their mitochondrial targeting in the NATIVEpr-GFP library when targeted by their native MTS. Although some proteins in this library could not be visualized owing to low expression levels, we found ten new mitochondrial proteins, most of which are encoded by genes without an annotated function or name (Fig. 3b).

Notably, deletion of the MTS from all MTS-containing proteins (by using a *TEF2pr-mCherry* cassette without an MTS) uncovered five proteins that robustly localized to mitochondria even in the absence of their predicted MTS (Supplementary Fig. 3c). We investigated whether the MTS-truncated versions of two of them, Tam41 and Coq2, were still translocated into mitochondria using in vitro translocation assays (Supplementary Fig. 5). Indeed, Tam41 was efficiently imported into mitochondria even when its MTS was truncated. A TargetP analysis of Tam41 for internal MTS-like signals (iMTS)[25] revealed a C′ iMTS that could be involved in this process (Supplementary Fig. 5a). In contrast, Coq2 lacking its MTS, which was targeted to the mitochondrial membrane (Supplementary Fig. 3c), could no longer be imported in vitro into isolated organelles (Supplementary Fig. 5b), which suggests that for this protein, targeting and translocation information are found in distinct regions.

When we looked at all of the mitochondrial proteins visualized to date with a GFP tag (635 proteins as annotated by the C′ GFP, N′ NATIVEpr-MTS-GFP, or NOP1pr-GFP libraries), it appeared that 70 such proteins have neither a predicted MTS nor a transmembrane domain (TMD) that might help in their targeting to mitochondria (Supplementary Fig. 3d). Although six of these proteins rely on the MIA complex (through a Cx(9)C cysteine-rich domain)[26] and four rely on the SAM/TOB complex (through a β-barrel domain)[27] for their targeting and translocation to mitochondria, the rest might target to mitochondria using other, as yet uncharacterized signals.

Several N′ proteins that localized to mitochondria also showed localization to another organelle in the same cell or were localized differently when C′-tagged. This suggests that some of these proteins are dually targeted[28]. We imaged a subset of these proteins from the NATIVEpr-GFP library under several growth conditions (Supplementary Fig. 6) and found that they could indeed reside in a variety of organelles depending on the medium. Knowledge on the dynamics of such mitochondrial proteins may improve understanding of the cross-talk of mitochondria with other cellular compartments[29,30].

Protein-fragment complementation SWAT libraries can be used for systematic measurement of protein–protein interactions. We next used our SWAT parental library to build four libraries for assaying protein–protein interactions. We based our new libraries on two protein-fragment complementation assay (PCA) approaches: split Venus[31] and the split dihydrofolate reductase (DHFR) enzyme[4].

The DHFR PCA reporter confers resistance to the cytostatic drug methotrexate, allowing growth when the two proteins tagged with the two fragments of DHFR interact (the N′ fragment, termed

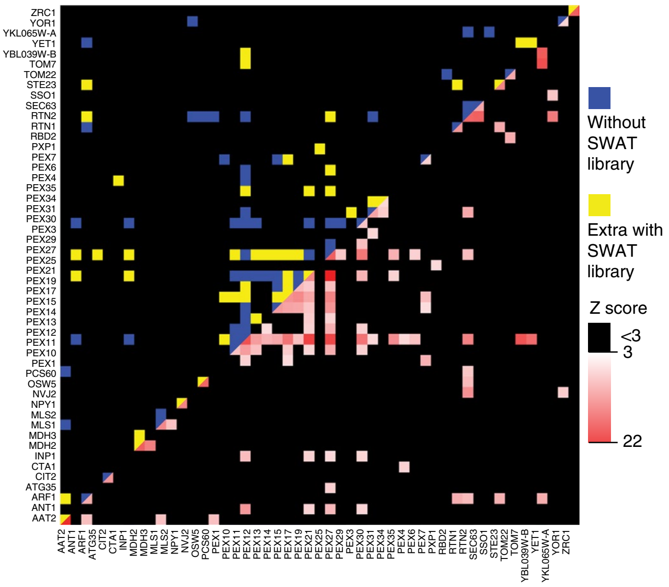


Fig. 4 | Protein-fragment complementation libraries enable systematic analysis of protein–protein interactions. Peroxisome DHFR PCA. Four yeast strain arrays were constructed, each representing 89 peroxisome-related genes tagged with either DHFR F[1,2] or DHFR F[3] fragments of methotrexate-resistant DHFR at either their N or C terminus. These strains were mated to test every protein pair in all permutations. An interaction between two proteins brings the DHFR fragments together, resulting in their folding, reconstitution of activity, and growth of strains in the presence of methotrexate. Blue squares represent interactions discovered by the C′-tagged strains alone. Yellow squares represent interactions that required at least one N′-tagged strain for their discovery. The white-to-red-spectrum squares correspond to the *Z* scores for the interaction. Only proteins with at least one interaction are depicted. A complete list of *Z* scores for the interaction is presented in Supplementary Table 4.

DHFR F[1,2], and the C′ fragment, termed DHFR F[3]). A previous large-scale protein interactome was determined in yeast with two C′-tagged DHFR fragment libraries[4]. As the C′ tag does not always enable the correct localization of proteins, and because interaction between the two fragments requires a specific topology of membrane proteins (the two fragments have to be facing the same side of the membrane to enable the enzymatic function to be reconstituted), we wished to investigate how complementary N′ DHFR PCA libraries could improve the coverage of protein–protein interactions. We made two new libraries: NATIVEpr-DHFR-F[1,2] and NATIVEpr-DHFR-F[3] (Fig. 1b). As a test case of these libraries, we focused on peroxisome proteins to generate a whole-organelle interactome. We used strains of peroxisomal proteins from all four DHFR libraries (the two previously published C′ ones and both new N′ ones). Strains from all four libraries were used for a pairwise DHFR PCA screen to test for interactions among the 89 peroxisome-related proteins (each library was mated with the two others from the opposing mating type and assayed).

The DHFR PCA revealed 230 positive results (Fig. 4 and Supplementary Table 4). We examined reproducibility by comparing the two strains with the same tagged proteins and the same position of the DHFR fragments (N′ or C′), but with the DHFR fragments swapped (for example, X-DHFR-F[1,2] + DHFR-F[3]-Y compared with X-DHFR-F[3] + DHFR-F[1,2]-Y). From 165 that had such a paired setup, 120 were reproducible, which put the reproducibility at 73%. As a result of the paired setup, the number of unique interactions was 109. Of these, 55.9% have been reported in the literature, including many known complexes on the peroxisome membrane[32], whereas only 4.7% of the non-interactions were previously reported, thus supporting the validity of our scoring system.

Indeed, it seems as if having both the C′ and N′ libraries is important for grasping the entire interactome, as of the 109 unique interactions, 48 required the inclusion of N′ strains (Fig. 4) and 75 would not be found without the presence of the C′ strains. Unfortunately, the majority of proteins residing in the peroxisome matrix did not show any interaction in either library, which suggests that DHFR substrate availability or DHFR reconstitution efficiency is too limited in the peroxisome.

In addition, we built two libraries based on the Venus PCA. In these libraries, two complementary N- and C-terminal fragments of the fluorescent protein Venus (YFP) (the C′ fragment termed VC and N′ fragment termed VN) were N′-fused to all proteins in opposite yeast mating types (Fig. 2a). Simple crossing of the two mating type strains resulted in diploids that were used to detect protein association by reconstitution of the full Venus fluorescence. Using these libraries, we repeated the whole-peroxisomal interactome (Supplementary Table 5) and found that, in general, this approach was less specific than the DHFR PCA. However, we did find several high-confidence, newly predicted interactions, such as the one between Inp1 and Pex17, that we were able to verify by yeast two-hybrid (Y2H) assay (Supplementary Fig. 7).

Split Venus libraries can be used to assay N′ topology. The strength of the Venus PCA library is that, as a result of the intrinsic affinity of the Venus fragments, it can be used to study membrane topology in addition to protein–protein interactions (Fig. 5a). To do this, we used the N′-tagged library including the C′ fragment of the split Venus cassette (VC) under the *TEF2* constitutive promoter, termed *TEF2pr-VC* (excluding SP- and MTS-bearing proteins), and mated it with a strain containing the N′ half of the split Venus cassette (VN) not conjugated to any other protein and therefore freely distributed in the cytosol (termed cyto-VN). This configuration should allow complementation of the VC and VN fragments, which results in a fluorescent signal, only if the N′ end of the VC-tagged protein faces the cytosol where the VN is prevalent (Fig. 5a). After selection for diploids, the fluorescence of each strain was quantified and localization was assigned by means of fluorescence microscopy (Supplementary Table 1). Cells that had fluorescence above a threshold level were termed N′ 'in' (facing the cytosol) (because the topology of 'out' could also result from a technical error giving lack of signal, this assignment could not be made unequivocally) (Supplementary Table 6). We verified this assignment for one protein, Scm4 (Supplementary Fig. 8). As a more general quality control step, we compared the abundance of proteins in the TEF2pr-mCherry library with the signal intensities that we measured in this complementation assay, as both libraries use the same promoter and should give a similar intensity profile. Indeed, the intensity of proteins that showed an in signal had a 0.82 two-sided Spearman correlation score with their TEF2pr-mCherry counterparts (Fig. 5b). For a subset of proteins in which the orientation of the C′ had been experimentally verified[33], we were able to use our topology predictions to also resolve TMD number (Supplementary Fig. 9). Currently, our method can clearly define only proteins whose N′ faces the cytosol. By anchoring the complementary Venus fragment in the lumen of organelles, it is possible to extend our method to define the topology of proteins whose N′ faces the interior of their respective organelles.

Discussion

Our new N′ tag genome-wide libraries enabled us to explore the proteome on several levels: abundance, localization, topology, and protein–protein interaction. We hope that the new information introduced here for uncharacterized proteins together with the

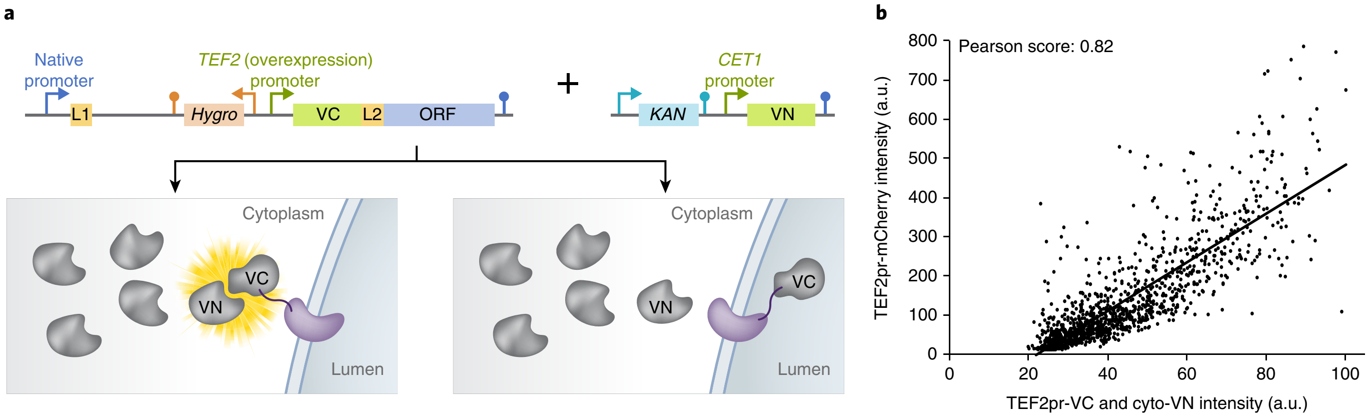


Fig. 5 | Protein-fragment complementation libraries enable systematic analysis of membrane protein topology. a, Scheme of split Venus analysis to determine topology for the N′ of proteins. The N′ TEF2pr-VC library was mated with a strain containing a cytosolic VN, and only if the N′ of a membrane protein faced the cytosol did complementation occur and a fluorescent signal appear, suggesting the topology of the protein's N terminus. A complete list of topology assignments is presented in Supplementary Table 6. **b**, Correlation between the TEF2pr-mCherry library and the TEF2pr-VC library with a cytosolic-VN (only 'in' assignments). Correlation score is two-sided Spearman correlation. Quantitation of abundance was performed once. Strains with a final abundance score less than 1 were excluded from the data analysis and Spearman correlation tests.

presence of these genes in our new libraries will promote the investigation of their functions.

Currently, our library is intended for use in an arrayed format. However, pooled experiments may be valuable in the future. For such cases, in principle, our pooled approach for sequencing the strains, which relies on the sequence of the L2 linker followed by the gene sequence, could serve as a pseudo-barcode, but would have to be developed into a quantitative assay[34]. An alternative and easy strategy for using SWAT-derived libraries in a pooled fashion is mating them with a barcoder library[35].

The parental N′ SWAT library with its easy-to-use swapping ability enables endless new possibilities for array-wide protein investigation. In a short time period and at a fraction of the cost incurred to date with other approaches, any yeast laboratory can make its very own library harboring a variety of selection markers, promoters, UTRs, targeting signals, fluorophores, affinity tags, or any other genetic element of choice. With this platform, the systematic exploration of any protein is no longer restricted and can be done with either N′ or C′ tagging[7]. Together, these approaches should contribute greatly to knowledge on the workings of living cells.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41592-018-0044-9.

Received: 21 December 2017; Accepted: 10 May 2018;
Published online: 09 July 2018

References

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. Huh, W.-K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
4. Tarassov, K. et al. An in vivo map of the yeast protein interactome. *Science* **320**, 1465–1470 (2008).
5. Yofe, I. et al. One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. *Nat. Methods* **13**, 371–378 (2016).
6. Khmelinskii, A., Meurer, M., Duishoev, N., Delhomme, N. & Knop, M. Seamless gene tagging by endonuclease-driven homologous recombination. *PLoS One* **6**, e23794 (2011).
7. Meurer, M. et al. A genome-wide resource for high-throughput genomic tagging of yeast ORFs. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2017/11/30/226811 (2017).
8. Engel, S. R. & Cherry, J. M. The genomic new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the Saccharomyces Genome Database. *Database (Oxf.)* **2013**, bat012 (2013).
9. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
10. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119–122 (1995).
11. Sun, J. et al. Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **109**, 2082–2092 (2012).
12. 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).
13. Picotti, P. et al. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266–270 (2013).
14. Weinberg, D. E. et al. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* **14**, 1787–1799 (2016).
15. 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).
16. Neymotin, B., Athanasiadou, R. & Gresham, D. Determination of in vivo RNA kinetics using RATE-seq. *RNA* **20**, 1645–1652 (2014).
17. 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).
18. Chen, X. & Zhang, J. The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst.* **2**, 347–354 (2016).
19. Weill, U. et al. Toolbox: creating a systematic database of secretory pathway proteins uncovers new cargo for COPI. *Traffic* **19**, 370–379 (2018).
20. Morgenstern, M. et al. Definition of a high-confidence mitochondrial proteome at quantitative scale. *Cell Rep.* **19**, 2836–2852 (2017).
21. Fukasawa, Y. et al. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteom.* **14**, 1113–1126 (2015).
22. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).
23. Vögtle, F.-N. et al. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **139**, 428–439 (2009).
24. Venne, A. S., Vögtle, F.-N., Meisinger, C., Sickmann, A. & Zahedi, R. P. Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55. *J. Proteome Res.* **12**, 3823–3830 (2013).

25. Backes, S. et al. Tom70 enhances mitochondrial preprotein import efficiency by binding to internal targeting sequences. *J. Cell Biol.* **217**, 1369–1382 (2018).
26. Chacinska, A. et al. Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. *EMBO J.* **23**, 3735–3746 (2004).
27. Wiedemann, N. et al. Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* **424**, 565–571 (2003).
28. Ben-Menachem, R. & Pines, O. Detection of dual targeting and dual function of mitochondrial proteins in yeast. in *Mitochondria: Practical Protocols* (eds. Mokranjac, D. & Perocchi, F.) 179–195 (Springer, New York, 2017).
29. Eisenberg-Bord, M. & Schuldiner, M. Mitochatting: if only we could be a fly on the cell wall. *Biochim. Biophys. Acta* **1864**, 1469–1480 (2017).
30. Eisenberg-Bord, M. & Schuldiner, M. Ground control to major TOM: mitochondria-nucleus communication. *FEBS J.* **284**, 196–210 (2017).
31. Jin, L. et al. Random insertion of split-cans of the fluorescent protein Venus into Shaker channels yields voltage sensitive probes with improved membrane localization in mammalian cells. *J. Neurosci. Methods* **199**, 1–9 (2011).
32. Erdmann, R. Assembly, maintenance and dynamics of peroxisomes. *Biochim. Biophys. Acta* **1863**, 787–789 (2016).
33. Kim, H., Melén, K., Osterberg, M. & von Heijne, G. A global topology map of the *Saccharomyces cerevisiae* membrane proteome. *Proc. Natl. Acad. Sci. USA* **103**, 11142–11147 (2006).
34. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).
35. Douglas, A. C. et al. Functional analysis with a barcoder yeast gene overexpression system. *G3 (Bethesda)* **2**, 1279–1289 (2012).

Acknowledgements

We thank Y. Peleg for plasmid construction, G. Brodsky for graphics, R. Rotkopf for support in statistical analysis, K. Tedrick for technical help with the Y2H experiments, and C. Meisinger and N. Vögtle for help with the MTS assignments. We thank G. Krieger for helpful discussions and technical help. We thank C. Ungermann (University of Osnabrück, Germany) and W.-K. Huh (Seoul National University, Seoul, South Korea) for plasmids. The work in the Schuldiner laboratory was supported by ERC CoG Peroxisystem (646604), SFB 1190 from the DFG, a Mitzutani foundation grant, and a VolksWagen foundation grant (93092). The collaborative work on this manuscript done by the Schuldiner, Pines, Herrmann, and Rapaport laboratories was supported by a DIP grant (P17516). Work at the Rachubinski lab was supported by Foundation Grant FDN-143289 from the Canadian Institutes of Health Research. Work in the Michnick lab was supported by Canadian Institutes of Health Research grant MOP-GMX-152556. Work in the Levy lab was supported by Israel Science Foundation grants 1775/12 and 2179/14. U.W. and D.D. are recipients of the Azrieli student-award grant. M.S. is an Incumbent of the Dr. Gilbert Omenn and Martha Darling Professorial Chair in Molecular Genetics.

Author contributions

U.W., I.Y., and M.S. conceived the study. U.W., I.Y., E.S., B.S., D.D., J.N., R.B.M., Z.A., O.G., N.H., S.C., K.K., B.K., J.L., F.B., J.K., and S.B.-D. carried out the investigation. M.S. and U.W. wrote the manuscript. All of the authors reviewed and edited the manuscript. E.Z., J.M.H., R.A.R., O.P., D.R., S.W.M., E.D.L., and M.S. supervised the work and acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0044-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Plasmid construction. We constructed plasmids using restriction-free cloning methods[36]. For a complete list of plasmids, see Supplementary Table 7. The I-SceI restriction site sequence was agttacgctagggataacagggtaatatag. The protein linker sequences (which also served as the generic recombination sites) were as follows: L1, 5′-cgtacgctgcaggtcgacggtggcggttctggcggtggcggatcc-3′; L2, 5′-ggcggttcctctggtggtggtggtgcgacagagaattcatcgatg-3′. Underlined sequences are the primer sequences used for amplification of the tagging module, corresponding to the pYM series sequences (S1and S4, respectively)[6]. The use of these sequences ensured compatibility with existing oligo collections for these popular module sets.

The tagging modules included the constitutive promoter of the *SpNOP1* gene[6] 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 GFP used in the cassettes of both the *NOP1pr-GFP* and *NATIVEpr-GFP* libraries is Superfolder GFP[9].

The Kar2 SP sequence used in the *SWAT-SP-GFP* module was atgtttttcaacagactaagcgctggcaagctgctggtaccactctccgtggtcctgtacgcccttttcgtggtaatattacctttacagaattctttccactcctccaatgttttagttagaggtgccgat.

The codon-modified (to avoid altered recombination) Kar2 SP sequence used in the donor NAT::TEF2pr-SPKar2-mCherry plasmid was atgttcttcaatagattgtcagctgggaagcttcttgtgccactgtctgtagttctttacgcactgttcgtagtgatactacccctgcaaaactcctttcactcttctaatgtcctggtcagaggcgcagac.

The MTS of *Neurospora crassa* OR74A ATP synthase protein 9 sequence used in the *SWAT-MTS-GFP* module was atggcctccactcgtgtcctcgcctctcgcctggcctcccggatggctgcttccgccaaggttgcccgccctgctgtccgcgttgctcaggtcagcaagcgcaccatccagactggctccccctccagaccctcaagcgcacccagatgacctccatcgtcaacgccaccacccgccaggctttccagaagcgcgcctac.

The codon-modified (to avoid altered recombination) MTS of *N. crassa* OR74A ATP synthase protein 9 sequence used in the donor NAT::*TEF2pr-MTS-mCherry* plasmid was atggcttctaccagagttttggcttctagattggcttctagaatggcagctagtgctaaggttgctagaccagctgttagagttgcacaagtttctaagagaacaatacaaaccggttctccattgcaaaccttgaagagaacccaaatgacttctatcgttaacgctactaccagacaagcatttcaaaagagagcttac.

The ORFs of Tam41 and Coq2 versions lacking the N-terminal 28 or 35 amino acids were amplified and cloned into the SacI and SalI or SacI and HindIII sites (respectively) of pGEM4 (Promega).

Primer choice and design. Total number of genes annotated in SGD currently stands at 6,075, excluding dubious ones. Because of the structure of our tagging cassette, we did not attempt to tag the 62 yeast proteins that include an N′ intron or the 250 genes that have identical homologs in the genome. Of the remaining 5,763 genes that can accurately be tagged with our N′ SWAT cassette, we had already attempted to use 1,847 in our previous work[5]. For the rest we designed primers and attempted to create them.

Primers for amplification of transformation cassettes and gene-specific targeting were designed with the Primers-4-Yeast web tool[37] (http://wws.weizmann.ac.il/Primers-4-Yeast) using the pYM plasmid type[36]. 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, as described in the Primers-4-Yeast web tool. For N′ tagging of SP or MTS containing proteins, homology sequences were designed to insert the cassette five amino acids downstream from the predicted cleavage point. Primers for validation of tagging 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 8.

High-throughput yeast transformations. The BY4741 laboratory strain[38], which is the basis for most systematic yeast libraries, was used as the master strain for the collection. The SWAT-GFP, SWAT-MTS-GFP and SWAT-SP-GFP acceptor modules (Supplementary Table 7) 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[39] 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 and then 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. To select pristine strains to be included in the final library, we picked four clones for each gene, and performed several quality control steps. Transformations that failed to yield four clones were repeated (462), and those that still failed were redone using resynthesized primer pairs (199). After these efforts, we obtained coverage of 95% of the anticipated yeast genes. Out of the 251 proteins that could not be tagged, 33 have an overexpression growth inhibition[40], 77 are essential proteins[2], and 47 proteins showed GO term enrichment for "cytoplasmic translation" ($P = 0.016448$)[8] (Supplementary Table 9).

The quality control that each strain underwent was as follows: (i) validation of integration locus by PCR (Supplementary Table 3), performed 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 (Supplementary Table 8). Strains of proteins that have a signal peptide did not undergo a PCR check. Strains that did not have a positive PCR for any of the four clones were not included in the final clone library (251 proteins; this could mean that there were not clones obtained or that no clone gave a positive PCR). (ii) Detection of the fusion protein by fluorescence microscopy: two clones were imaged by fluorescence microscopy, and we reviewed images manually to assign up to three localizations to each clone. Strains with a nondistinct pattern were given the assignment of 'ambiguous'. Strains with fluorescent signal that was quantified as below background signal were given the assignment of 'below threshold'. Assignment categories were bud, bud neck, cell periphery, cytosol, ER, mitochondria, nuclear periphery, nucleolus, nucleus, punctate, vacuole and vacuole membrane. (iii) Determination of the SWAT swapping capability (see below). (iv) Sequencing (Anchor-seq) to ensure correct reading frame: we used a targeted-sequencing strategy detailed in ref. [7] to verify the junction encompassing the 3′ end of the cassette and the 5′ end of each gene. Briefly, we pooled all strains from the SWAT library, extracted their genomic DNA, and sheared it into fragments of 300–800 bp that were gel-purified, ligated to Anchor-seq adaptors and subjected to two rounds of PCR. The sequences of adaptors and of oligonucleotides are detailed in Supplementary Table 8. This protocol enriched specifically the junctions of interest, which we then sequenced by next-generation sequencing. We subsequently analyzed the reads to classify each ORF into one of three categories corresponding to validated sequences, sequences containing a frameshift, or sequences containing a point mutation (Supplementary Table 3). Sequencing was performed in two steps; first sequencing was performed on all four initial clones. In this round (Anchor-seq round 1; Supplementary Table 3) the reads were 150 bp from the L2 linker into the coding sequence. Analysis was performed to determine whether strains were 'positive', meaning that the cassette was inserted correctly and no mismatches or indels were observed; 'not detected', meaning that the sequence could not be found after sequencing; 'mismatches', meaning that either a mismatch or an indel was found; or 'half', meaning that there was an imbalance in read count between the first and second halves of the read. To complete the analysis also for the previously published strains[5], were also sequenced those. The reads were again 150 bp from the L2 linker into the coding sequence. The analysis was done as above and we added information of 'bad linkers', meaning there was only a mismatch in the L2 linker.

Finally, after the final clones were selected for the full genome SWAT library, a second round of sequencing was carried out (Anchor-seq round 2; Supplementary Table 3). This time the reads were only 92 bp from the L2 linker into the coding sequence. Analysis was done as above but differentiated between 'mismatch', which means that some bases did not match the expected sequence, and 'indel', which means that an insertion or a deletion occurred within the sequence. We also annotated 'low read count', which means that the sequence was correct but observed fewer times than would be expected, and we added information on the amount of base pairs altered in the mismatch or indel strains.

A strain was removed from the library if the second round of Anchor-seq showed that it had an indel, or if it could not be detected in the second round of Anchor-seq and the first round of Anchor-seq showed it to have a mismatch. Proteins bearing a signal peptide that did not undergo a PCR check were removed if they were not annotated as positive in the two Anchor-seq rounds. All other alterations from expected sequence are highlighted as remarks in Supplementary Table 3.

Strains with a validated sequence, consistent localization assignment, and swapping capacity and that had been validated by PCR were chosen to compose the final SWAT full-genome library that contained 5,457 strains (Supplementary Tables 1 and 3). The rigorous quality control should maximize the utility of this parental library, which will become a basis for multiple future N′ yeast full-genome libraries.

Donor strain construction. Donor strains were constructed on the background of an SGA[41] 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 *Kluyveromyces lactis URA3* selection marker into the *can1*Δ locus, upstream of the *STE2pr-SpHIS5* fragment (used for selection of MATa). A *Gal1pr-I-SceI* fragment was then introduced to replace the *URA3* selection, resulting in *can1*Δ*::GAL1pr-SceI::STE2pr-SpHIS5* (strain yMS2085).

Analysis of swapping-procedure efficiency. First, to see whether the native promoter/regulation GFP swap can restore the regulation of the native promoter, we swapped three conditionally induced genes. *NOP1pr-GFP-GAL2* and *NATIVEpr-GFP-GAL2* were grown for 4 h in either liquid SD (2% glucose) or SG (2% galactose) medium. *NOP1pr-GFP-SUC2* and *NATIVEpr-GFP-SUC2* were grown for 4 h in either liquid SD glucose or synthetic medium with no glucose. *NOP1pr-GFP-PHO5* and *NATIVEpr-GFP-PHO5* were grown for 4 h in either liquid SD complete or SD-phosphate medium (Supplementary Fig. 1). Next, we measured the swapping efficiency of *NOP1pr-GFP* strains to *TEF2pr-mCherry* for all four clones. The clones were imaged via a high-content screening platform in brightfield, GFP and Cherry channels. We reviewed images of all clones manually and assigned up to three localizations to each clone. Assignment categories were as above. Preference for inclusion in the SWAT-GFP library was given to clones that showed a similar localization to the *NOP1pr-GFP* tag.

Automated manipulation of yeast libraries. We conducted automated strain maintenance and manipulation using a RoToR benchtop colony arrayer[42] (Singer Instruments). We carried out SGA procedures[41] for mating of the parental SWAT-GFP collections with donor strains bearing the native promoter/regulation GFP donor (Supplementary Table 7; pSD-N9), the NAT:*TEF2pr-mCherry* donor (Supplementary Table 7; pSD-N15/16/21), the HYGRO:*TEF2pr-VC* donor (Supplementary Table 7; pSD-23), and the KAN:*CET1pr-VN* donor (Supplementary Table 7; pSD-N24). After double-mutant selection, all libraries were selected for MATα haploids except for the *CET1pr-VN* library, which was selected for MATa. Then tag swapping was prompted by growth on yeast extract peptone (YEP)-galactose (2%) media for 1–2 d to induce I-SceI expression. Tag swapping was then selected by two cycles of growth over night on SD + 5-FOA (1 g/L) media for *NATIVEpr-GFP* library, yeast extract peptone dextrose (YEPD) + nourseothricin (NAT; 200 μg/ml) for the *TEF2pr-mCherry* library, SD + 5-FOA (1 g/L) + hygromycin B (200 μg/ml) for the *TEF2pr-VC* library, or SD + 5-FOA (1 g/L) + g418 (200 μg/ml) for the *CET1pr-VN* library.

High-throughput microscopy. We carried out high-content screening of strain collections using an automated microscopy setup (ScanR system; Olympus) as previously described[12]. 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), BFP (excitation, 402/15 nm; emission, 455/50 nm) and brightfield channels. Images were analyzed using the ScanR Analysis software 2.7.0 (r3429) x64 (Olympus), and single cells were recognized on the basis of the brightfield channel. Measures of cell size, shape and fluorescence signals were extracted. For localization assignments, we reviewed images manually using ImageJ (1.51p Java1.8.0_144 (64-bit)). 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 includes structures such as the Golgi apparatus, peroxisomes, endosomes, p-bodies, inclusions, lipid droplets, other vesicular structures and subdomain compartments) (Supplementary Table 1). All images of the N′ library strains can be found and downloaded at our Loqate database (http://www.weizmann.ac.il/molgen/loqate).

Computational quantification of single-cell fluorophore intensity. We measured the median GFP/mCherry intensity for each strain using single-cell recognition software (scanR Analysis software 2.7.0 (r3429) x64; Olympus) as previously described[12]. Strains with fewer than 30 recognized cells were excluded. We obtained the baseline autofluorescence level of each plate from strains not expressing GFP. We then calculated each strain's final score by subtraction of this value from each strain median GFP/mCherry intensity. Strains with a final score less than 1 were excluded from the data analysis and two-sided Spearman correlation tests that were performed with R studio version 0.99.486.

Data processing. We compared the subcellular-localization annotations of the NOP1pr-GFP, NOP1pr-MTS-GFP, and NOP1pr-SP-GFP libraries with those of the C′-tag library (comprising data from two previously published datasets[3,12]). 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. 'N′ only' was assigned if the C′ localization was classified as below threshold or ambiguous, or if no assignment existed. 'C′ only' was assigned if the N′ localization was classified as below threshold or ambiguous, or if no assignment existed. 'Neither tag' was assigned if both the N′ localization and C′ localization were classified as below threshold or ambiguous, or if no assignment existed. All other cases were classified as 'different' (Fig. 3a). All of the calculations were performed with Python 2.7 software.

Mitochondrial targeting signal predictions of yeast proteins. We compiled MTS determinations for all yeast proteins based on both experimental evidence (EE) from previous studies (EE1[23], EE2[24] and MTS prediction algorithms (Mitofates[21] and TargetP version 1.1[22])) (Supplementary Table 2). As each of the four sources support different MTS designations, we employed a scoring method for the analysis results whereby a prediction is accounted for if it complies with the following rules:

1. An MTS must be > 6 amino acids in length based on experimental evidence or prediction.
2. Any protein identified in EE1 five times or more.
3. Any protein that Mitofates predicted as MTS-containing with a score of > 0.5 (reported precision of 0.83).
4. A few cases were manually assigned for known MTS-containing proteins from literature review.
5. Bona fide nonmitochondrial proteins were removed after manual review.

Using these criteria, we designated 420 proteins as having MTS. Then, the MTS cleavage site (distance in amino acids from N′) was selected by the experimental evidence and predictions. Selection of the cleavage site was done by the following hierarchy:

1. Site was consistent (that is, within 5 amino acids apart) between EE1 and EE2.
2. Site was consistent between EE1 and Mitofates prediction.
3. Site was consistent between EE1 and TargetP prediction.
4. Site was consistent between EE2 and Mitofates prediction.
5. Site was consistent between EE2 and TargetP prediction.
6. When there was no consistency between experimental evidence, the site was first determined by EE1, and only if not available was it then determined by EE2.
7. For MTSs classified only by predictions, a site was given priority if it was consistent between Mitofates prediction and TargetP prediction. If no consistency was found, then the site was chosen as determined by Mitofates.
8. For a few cases the cleavage site was picked manually on the basis of previous evidence.

Subcellular fractionation and western blotting analysis (Supplementary Fig. 4). Yeast cultures were grown to an A_{600} of 1.5. Mitochondria were isolated as described previously[43]. Spheroplasts were prepared in the presence of zymolyase 20 T (MP Biomedicals, Irvine, CA). Equivalent portions from fractions of the total (T), cytosol (C) and mitochondria (M) were analyzed by western blotting using α α to follow the tagged protein (either Ysa1 or Kgd2), αHsp60 as a mitochondrial marker and αHxk1 as a cytosolic marker. Full scans of all blots are shown in Supplementary Fig. 10a–c.

Import of radiolabeled proteins into isolated mitochondria (Supplementary Fig. 5). Isolation of yeast mitochondria and import reactions were essentially performed as described previously[44] in the following import buffer: 500 mM sorbitol, 50 mM Hepes, pH 7.4, 80 mM KCl, 10 mM magnesium acetate, and 2 mM KH_2PO_4. Mitochondria were energized by the addition of 2 mM ATP and 2 mM NADH before radiolabeled precursor proteins were added. To dissipate the membrane potential, a mixture of 1 μg/ml valinomycin, 8.8 μg/ml antimycin, and 17 μg/ml oligomycin was added to the mitochondria. Precursor proteins were incubated with mitochondria for different times at 25 °C before non-imported protein was degraded by the addition of 100 μg/ml proteinase K. Full scans of all blots are shown in Supplementary Fig. 10d,e.

Mitochondrial protein dual-localization analysis (Supplementary Fig. 6). Strains showing mitochondrial dual localization in the N′ *NATIVEpr-GFP* library (Supplementary Table 1) were arrayed into liquid 96-well polystyrene growth plates. Liquid cultures were grown overnight in SD medium at 30 °C. Cells were back-diluted to ~0.25 OD_{600} into four plates, each containing a different medium: glucose 2%, glycerol 2%, galactose 2% or glucose 0.2%. Plates were then grown for 4 h at 30 °C to reach logarithmic growth phase. Strains from all four plates in addition to the original overnight plate were transferred into glass-bottom 384-well microscope plates (Matrical Bioscience) coated with concanavalin A (Sigma-Aldrich) to allow cell adhesion. Wells were washed twice in appropriate medium to remove floating cells and reach cell monolayer. Manual microscopy was performed with the VisiScope Confocal Cell Explorer system, composed of a Zeiss Yokogawa spinning disk scanning unit (CSU-W1) coupled with an inverted Olympus IX83 microscope. Images were acquired using a 60× oil lens and captured by a connected PCO-Edge sCMOS (scientific complementary metal-oxide semiconductor) camera, controlled by VisView software, with a wavelength of 488 nm (GFP). Images were transferred to ImageJ (1.51p Java1.8.0_144 (64-bit)) for slight, linear adjustments to contrast and brightness.

SWAT DHFR PCA library construction. Antibiotic resistance genes *nat1* and *hph* were PCR-amplified respectively from plasmids pAG25 and pAG32[45] with primers DHFR-F1 and DHFR-F2 (Supplementary Table 8). Strain yMS2085[5] was transformed with the PCR products to create strains yMS2085-NAT1 and yMS2085-HPH, where each resistance gene is integrated on chromosome V between genes *CAJ1* and *TPA1*[46]. Next, SWAT DHFR PCA donor plasmids were created. First, DHFR F[1,2] and DHFR F[3] were PCR-amplified respectively from pAG25-linker-DHFR F[1,2] and pAG32-linker-DHFR F[3][4] with primer pairs DHFR-F2 and DHFR-R2, and DHFR-F3 and DHFR-R3 (Supplementary Table 8). The PCR products were cloned into pSD-N2 (Supplementary Table 7) with restriction enzymes BamHI and SpeI to form pSD-N25 and pSD-N26 (Supplementary Table 7). Strain yMS2085-NAT1 was transformed with pSD-N25 to create ySWAT-DHFR-F[1,2], and yMS2085-HPH was transformed with pSD-N26 to make ySWAT-DHFR-[F3].

The resulting strains were used as SWAT donor strains to generate two libraries with DHFR-F[1,2] or DHFR-F[3] tagged N-terminally of 89 peroxisome-related genes, according to the procedure previously described[5].

SWAT DHFR PCA screen. Four libraries were used for the DHFR PCA screen to test for interactions among the 89 peroxisome-related proteins. These libraries are the SWAT-based DHFR F[1,2] library (87 strains available out of 89), the

SWAT-based DHFR F[3] library (88 strains), a library with C-terminal DHFR F[1,2] tags (65 strains) and a library with C-terminal DHFR F[3] tags (75 strains)[4]. Each DHFR F[1,2] strain was mated with each of the DHFR F[3] strains by overnight incubation on YPD. The strains were organized in such a way that each row contained the same DHFR F[1,2] strain and each column the same DHFR F[3] strain, in a 1536-format on 24 plates in total. After mating, diploid cells were selected for by incubation for 2 d on YPD medium with 100 μg/ml nourseothricin (Werner Bioagents) and 250 μg/ml hygromycin B (Wisent Bioproducts). This step was repeated once. Next, the strains were transferred to synthetic complete medium (4% (w/v) Noble agar) with 200 μg/ml methotrexate (Bioshop Canada) and without adenine or ammonium sulfate. Pictures of the strains were taken at the start of the experiment and after 4 d of incubation at 30 °C. Every protein–protein interaction was tested in duplicate and all plate handling and imaging was done with a BioMatrix automated plate handler (S&P Robotics).

SWAT DHFR PCA data analysis. The integrated colony densities, which are an estimation of colony volume, were obtained with a custom-made ImageJ (1.51p Java1.8.0_144 (64-bit)) script that measures the integrated colony density by multiplication of the colony area with its mean intensity. To account for variation in the initial cell material deposited at the start of the experiment, the day 4 colony densities were corrected with their day 0 values through linear regression and normalization. Some crossed strains show higher overall background growth. To correct for this phenomenon, we subtracted the mean of the median row (identical F[1,2] strains) and median column (identical F[3] strains) values where the strain is located from the colony size. These two normalization steps improved the correlation between our results and those in the literature[47]. The colony sizes on each plate showed a normal distribution (Shapiro test, P values $< 10^{-32}$), and the Z scores of the colony sizes were calculated on the basis of the distribution within each plate. A result was considered positive if the minimum Z score (of the two duplicates) was greater than 3. From the positive results, 48% had been detected in previous protein–protein interaction studies (http://www.biogrid.org), not including results of the original C′ DHFR PCA screen[4]. However, this agreement dropped to less than 10% (4/46) when we considered results with relatively low Z scores (between 3 and 4.85) and at least one highly abundant protein. Therefore, those results with Z scores below 4.85 and at least one highly abundant protein were removed from the list of positive results. A highly abundant protein was defined as having a score above 30 according to GFP abundance data[3,5,12] with the GFP tagged in the same position as the DHFR PCA tag (N or C terminal). If these data were not available, GFP abundance values were taken from sources in which GFP was positioned on the other terminus. Calculations were done with R Studio version 0.99.486.

SWAT Venus PCA analysis (Supplementary Table 5). A yeast array of 92 strains, each expressing a peroxisomal associated protein (and some control strains), was compiled from the NOP1pr-GFP library (Supplementary Table 5). Strain manipulation was done on a RoToR benchtop colony arrayer[42] (Singer Instruments). We carried out SGA procedures[41] with donor strains bearing either the KAN:*CET1pr-VN* donor (Supplementary Table 7; pSD-N24) or the HYGRO:*TEF2pr-VC* donor (Supplementary Table 7; pSD-N23) with NAT:*PEX3-mCherry* as a peroxisomal marker. After double-mutant selection, the *TEF2pr-VC* array was selected for MATα haploids and the *CET1pr-VN* array was selected for MATa. Then tag swapping was prompted by growth on YEP-galactose (2%) media for 1–2 d to induce I-SceI expression. Tag swapping was then selected by two cycles of growth overnight on SD + 5-FOA (1 g/L) + hygromycin B (200 μg/ml) + nourseothricin (NAT; 200 μg/ml) for the *TEF2pr-VC* array, or SD + 5-FOA (1 g/L) + g418 (200 μg/ml) for the *CET1pr-VN* array. All strains from the two arrays were then crossed and selected for diploids. Strains were then imaged and analyzed for signal localization and intensity (see above) (Supplementary Table 5).

Yeast two-hybrid assay (Supplementary Fig. 7). The yeast strain HF7c[48] and protocols were from Clontech Laboratories. The full-length *PEX17* open reading frame was amplified from *S. cerevisiae* genomic DNA and inserted into the BamHI/SalI restriction sites of plasmids pGAD424 (AD) and pGBT9 (BD), respectively[49]. pGAD424-*INP1* and pGBT9-*INP1* plasmids have been described[50]. Plasmids were transformed into HF7c cells, and cells were cultured at 30 °C in synthetic dropout medium to an OD_{600} of 0.5–0.8 before being collected by centrifugation. The OD_{600} was adjusted to 1.0 for all cells, and 1 μl of a series of 1:10 dilutions (corresponding to an OD_{600} of 10^0, 10^{-1}, 10^{-2}, 10^{-3}) were then spotted onto selective plates, which were incubated at 30 °C for 3–5 d. -Leu -Trp medium selects for the presence of both pGAD424 and pGBT9 plasmids in cells, whereas -His -Leu -Trp medium selects for the presence of a protein–protein interaction.

TMD and N′ topology analysis. Transmembrane prediction was performed with the following programs using default parameters: TMHMM[51], HMMTOP[52], Phobius[53], Philius[54], and TOPCONS[55]. The TOPCONS results were taken for the TOPCONS algorithm itself, as well as all the component programs individually, Octopus, Polyphobius, Philius, Scampi and Spoctopus. As the results of Philius that were run separately and the results of Philius in the TOPCONS program were identical, only the results from the program run individually were used (Phobius and PolyPhobius gave different results). The annotation of TMD from the UniProt database was taken from the whole yeast proteome (accession number UP000002311) using the subcellular location "Transmembrane." Topology prediction was extracted from the same results of all the programs with the exception of UniProt, where there is no topology prediction. Custom scripts (available from the corresponding author on request) and manual analysis were used to parse the results.

Topology analysis using Venus PCA. Using automated strain maintenance and manipulation with a RoToR benchtop colony arrayer[42] (Singer Instruments), we carried out mating of the entire *TEF2pr-VC* library with a BY4741 strain containing HO::KAN-*CET1pr-VN*, NAT:*PEX3-mCherry* and the URA:*MTS-BFP* plasmid (Supplementary Table 7). After two rounds of diploid selection, strains were imaged and analyzed for signal localization and intensity (see above) (Supplementary Tables 1 and 6). Strains showing a signal above that of the *his3Δ1*::GFPdC control strain were considered as having the tagged proteins N′ facing the cytosol ('in').

Proteinase K protection assay (Supplementary Fig. 8). Mitochondria isolated from cells expressing tagged Scm4 were treated with the indicated amounts of proteinase K (PK) or trypsin. After inhibition of the proteases, samples were precipitated with trichloroacetic acid and analyzed by SDS–PAGE followed by immunodecoration with antibodies to the HA-tag or the relevant mitochondrial proteins. Tom70, a MOM protein exposed to the cytosol; Aco1 and Hep1, matrix proteins. Full scans of all blots are shown in Supplementary Fig. 10f,g.

Obtaining the libraries, plasmids, images and protocols. All strains, plasmids and libraries presented in this article are freely available from the corresponding author upon request. All protocols for using the SWAT strategy can be found on Protocol Exchange (https://www.nature.com/protocolexchange/labgroups/1106525) and on our lab website (http://www.weizmann.ac.il/molgen/Maya/SWAT). All images of the N′ library strains can be found and downloaded at our Loqate database (http://www.weizmann.ac.il/molgen/loqate).

Code availability. All original code used in this study is publicly available at https://github.com/uriweill/SWAT-N-scripts-.

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 author upon request.

References

36. 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).
37. Yofe, I. & Schuldiner, M. Primers-4-Yeast: a comprehensive web tool for planning primers for *Saccharomyces cerevisiae*. *Yeast* **31**, 77–80 (2014).
38. 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).
39. 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).
40. Sopko, R. et al. Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
41. Hin Yan Tong, A. & Boone, C. High-throughput strain construction and systematic synthetic lethal screening in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **36**, 1–19 (2007).
42. Cohen, Y. & Schuldiner, M. Advanced methods for high-throughput microscopy screening of genetically modified yeast libraries. *Methods Mol. Biol.* **781**, 127–159 (2011).
43. Knox, C., Sass, E., Neupert, W. & Pines, O. Import into mitochondria, folding and retrograde movement of fumarase in yeast. *J. Biol. Chem.* **273**, 25587–25593 (1998).
44. Weckbecker, D., Longen, S., Riemer, J. & Herrmann, J. M. Atp23 biogenesis reveals a chaperone-like folding activity of Mia40 in the IMS of mitochondria. *EMBO J.* **31**, 4348–4358 (2012).
45. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541–1553 (1999).
46. Flagfeldt, D. B., Siewers, V., Huang, L. & Nielsen, J. Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* **26**, 545–551 (2009).
47. Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).

48. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).

49. Bartel, P., Chien, C. T., Sternglanz, R. & Fields, S. Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* **14**, 920–924 (1993).

50. Knoblach, B. et al. An ER-peroxisome tether exerts peroxisome population control in yeast. *EMBO J.* **32**, 2439–2453 (2013).

51. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

52. Tusnády, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850 (2001).

53. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, 429–432 (2007).

54. 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).

55. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37**, 465–468 (2009).

nature research

Corresponding author(s): NMETH-BC33361C

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.

Sample size for SWAT full genome libraries include all recognized open reading frames of the Saccharomyces cerevisiae genome from: https://www.yeastgenome.org/
Since this is a systematic study, sample size was defined by the size of the yeast genome.

2. Data exclusions

Describe any data exclusions.

Strains with a final score intensity score lower than 1, were excluded from the data analysis and Spearman two sided correlation tests.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

In figures 2a-d and in figure5b quantification of protein abundance and localization based on microscopic imaging of yeast strains was performed once. In figure 3a-c imaging of yeast cells were performed once and figures represent entire field. In figure 4b two independent replicates were made for each protein-protein interaction assayed.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Since all verified open reading frames in the genome were included, no group devision was needed.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Since we did not aim to compare two or more populations, no blinding was necessary.

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.

R version 0.99.486 was used for the analysis done in figure 2a-c and figure 4b.
Python version 2.7 software was used for the analysis done in figure 2d.
ImageJ 1.51p Java1.8.0_144 (64-bit) was used to manually analyze images.
ScanR Analysis 2.7.0 (r3429) x64 was used to analyze images for signal intensity.
All original scripts used in this study are publicly available at:
https://github.com/uriweill/SWAT-N-scripts-.git

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.

No unique materials were used

9. Antibodies

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

α-HA: https://www.sigmaaldrich.com/catalog/product/roche/roahaha?lang=de&region=DE
Supplier name: Roche
Catalog number: 11867423001
Clone name: 3F10, monoclonal
Lot number: NA
Description of the validation of each primary antibody for the species and application:
Use Anti-HA High Affinity was used for the detection of native influenza hemagglutinin protein and recombinant proteins that contain the HA epitope using:
- Dot blots
- ELISA
- Immunocytochemistry
- Immunoprecipitation (https://www.sigmaaldrich.com/catalog/product/roche/roahaha?lang=en&region=IL#cited_3)
- Western blots (https://www.sigmaaldrich.com/catalog/product/roche/roahaha?lang=en&region=US#cited_1)(https://www.sigmaaldrich.com/catalog/product/roche/roahaha?lang=en&region=IL#cited_2)

Since Anti-HA High Affinity is a rat monoclonal, it is possible to use it in conjunction with murine monoclonals for double labeling.
α-HA dilution 1:2000

αTom70, α Aco1 and αHep1 were produced at the Rapaport lab:
All antibodies were raised in rabbit and diluted in 5% skim milk in TBS buffer in appropriate concentration. Proteins were detected by secondary goat antibody anti-rabbit (BIORAD) conjugated to horseradish peroxide (HRP).
α-Tom70 1:1000; α-Aco1 1:4000; α-Hep1 1:4000.
All antibodies were tested for specificity in lab by using the appropriate deletion strains as controls.

10. Eukaryotic cell lines

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

The BY4741 yeast laboratory strain (Brachmann et al., 1998) from the ATCC, which is the basis for most systematic yeast libraries, was used as the master strain for the collection.

b. Describe the method of cell line authentication used.

The yeast cell line used has not been a authenticated by us. It is a standard ATCC cell line.

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

Yeast cell lines do not get mycoplasma and hence were not tested for mycoplasma contamination

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.

No commonly misidentified cell lines were used

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 were 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.	Study did not include human research participants