Cell Reports
Resource

High-Resolution Mapping of Protein Concentration Reveals Principles of Proteome Architecture and Adaptation

Emmanuel D. Levy,[1,2,*] Jacqueline Kowarzyk,[1] and Stephen W. Michnick[1,*]
[1]Département de Biochimie and Centre Robert-Cedergren, Bio-Informatique et Génomique, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montreal, QC H3C 3J7, Canada
[2]Department of Structural Biology, Weizmann Institute of Science, Rehovot 7610001, Israel
*Correspondence: emmanuel.levy@weizmann.ac.il (E.D.L.), stephen.michnick@umontreal.ca (S.W.M.)
http://dx.doi.org/10.1016/j.celrep.2014.04.009

SUMMARY

A single yeast cell contains a hundred million protein molecules. How these proteins are organized to orchestrate living processes is a central question in biology. To probe this organization in vivo, we measured the local concentration of proteins based on the strength of their nonspecific interactions with a neutral reporter protein. We first used a cytosolic reporter and measured local concentrations for ~2,000 proteins in *S. cerevisiae*, with accuracy comparable to that of mass spectrometry. Localizing the reporter to membranes specifically increased the local concentration measured for membrane proteins. Comparing the concentrations measured by both reporters revealed that encounter frequencies between proteins are primarily dictated by their abundances. However, to change these encounter frequencies and restructure the proteome, as in adaptation, we find that changes in localization have more impact than changes in abundance. These results highlight how protein abundance and localization contribute to proteome organization and reorganization.

INTRODUCTION

Living cells rely on billions of molecules to form a self-organized system of remarkable complexity. A single cell of the budding yeast *S. cerevisiae* contains as many as a hundred million protein molecules and a mammalian cell about a hundred times more (Milo, 2013). Proteins are the main effectors of cellular functions, and understanding their organization and distribution in cells is a fundamental challenge. Hence, we would like to gain quantitative information on where, when, and with which partners a given protein interacts in the cell for the entire proteome.

S. cerevisiae is a powerful model organism to study protein properties on a large scale owing to the ease with which its genome can be manipulated. Fundamental properties of the *S. cerevisiae* proteome have been captured on a scale that includes practically all known gene products. Such properties include copy numbers of individual proteins and variations within individual cells (de Godoy et al., 2008; Ghaemmaghami et al., 2003; Newman et al., 2006), protein localization (Breker et al., 2013; Dénervaud et al., 2013; Huh et al., 2003; Tkach et al., 2012), and protein physical interactions (Breitkreutz et al., 2010; Gavin et al., 2006; Krogan et al., 2006; Tarassov et al., 2008; Uetz et al., 2000; Yu et al., 2008). These studies and subsequent work have demonstrated that protein concentrations are not uniform and instead follow a power law distribution spanning over four orders of magnitude (de Godoy et al., 2008; Futcher et al., 1999; Ghaemmaghami et al., 2003; Newman et al., 2006), meaning that a few high-copy proteins contribute to most of the total protein mass. Localization data showed that about half of all proteins are present in the cytosol, while the other half are in various subcellular compartments (Huh et al., 2003; Kumar et al., 2002). Protein-protein interaction data have revealed the widespread presence of complexes in the proteome (Gavin et al., 2006; Krogan et al., 2006; Rives and Galitski, 2003; Spirin and Mirny, 2003). It was also shown that interacting partners tend to be in the same subcellular compartment (Collins et al., 2007; Gavin et al., 2006; von Mering et al., 2002), exhibit similar abundance levels (Gavin et al., 2006), and show coexpression at the mRNA level (Collins et al., 2007; Gavin et al., 2006; Ge et al., 2001; Simonis et al., 2006; Tan et al., 2007). These observations are consistent with the concept of a cell being a well-organized system in which interacting components appear to be functionally coherent in time, space, and stoichiometry.

Importantly, however, such coherence is observed only for a subpopulation of all interactions (Gavin et al., 2006; Simonis et al., 2006). Indeed, coexpression patterns, colocalization, or similarity in stoichiometry are observed only in 23%–30%, 23%–55%, and 10%–22% of protein-protein interactions, respectively (Figure 1). In that respect, protein interactomes appear to have a cellular organization more akin to a social network. In such a network, interactions span a wide range of affinities (e.g., encounter, acquaintance, friend, family), which are expected to exhibit increasing levels of coherence. Similarly, proteins interact with one another on many different levels, from transient encounters to weak functional associations to obligate quaternary assemblies (Janin et al., 2008; Jones and Thornton, 1996). Any given protein may participate in any or all of these types of interaction, simultaneously or sequentially (Han et al.,

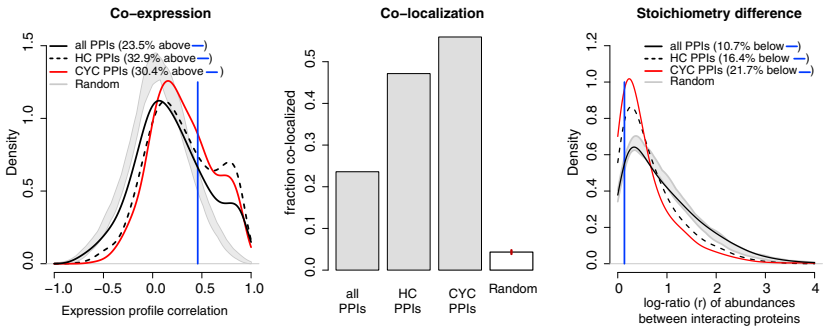

Figure 1. Coherence in Time, Space, and Stoichiometry between Interacting Proteins Is Statistically Significant but Limited in Magnitude

All binary protein-protein interactions (PPIs) were retrieved from BIOGRID (Stark et al., 2011), and the high-confidence (HC) PPIs are those supported by at least two publications. The CYC PPIs correspond to a matrix representation of CYC2008 (Pu et al., 2009) complexes with ten subunits or fewer. Coexpression data consist in all the conditions described in Gasch et al. (2000), on which pairwise Pearson correlation was computed between gene pairs. Colocalization was computed on the matrix provided in Data Set S1 but omitting all categories involving over 400 proteins. Stoichiometry differences were calculated by the absolute value of the log10 ratio of interacting protein abundances: $r = abs(\log10(AB^i / AB^j))$, where the two proteins considered have abundances AB^i and AB^j. If two subunits of a complex are present at identical levels in the cell, the ratio 'r' would thus be zero. Random values were calculated by shuffling protein pairs from the CYC2008 PPIs data. Shuffling was repeated 100 times, and the gray area (left and right panels) shows the minimum and maximum density observed at each value of the *x* axis. The blue line on the left panel delimits the upper 10^{th} percentile of the random distribution. We use it as the cutoff above which we consider coexpression to take place. The small red line (middle panel, on the "Random" bar) shows the SD of the random values observed. The blue line in the right panel shows the lower 10^{th} percentile of the random distribution. We use it as a cutoff below which we consider stoichiometry to be similar.

2004; McGuffee and Elcock, 2010). Such a fuzzy organization of the proteome is also consistent with observations that proteins can have many functions (Huberts and van der Klei, 2010). For example, although the ribosome is a highly dedicated and specialized molecular machine that translates mRNAs, many secondary functions, some of which are completely unrelated to translation, have been attributed to its constituent proteins (Warner and McIntosh, 2009).

To capture the intricacies of proteome organization, quantitative data regarding protein abundance, localization, and interactions are needed. Quantitative data are already available for protein abundance, but comparatively little quantitative data are available for protein localization and protein-protein interactions. For example, it is not known to what extent a membrane protein is isolated with respect to other compartments, i.e., if two cytosolic proteins encounter each other with a given frequency, how would this frequency change if one of the two proteins is now localized to the membrane? To address such questions, we devised a strategy to measure local protein concentrations in vivo on a proteome-wide scale, allowing us to probe global organizing principles of the *S. cerevisiae* proteome.

RESULTS

Measuring Local Protein Concentration through Nonspecific Interactions

Our strategy relied on the protein-fragment complementation assay (PCA) methodology (Figure 2A), which consists of fusing two complementary N- and C-terminal fragments of a reporter protein to two other proteins of interest. If the proteins of interest are in spatial proximity in the cell, then the two complementary fragments fold together, resulting in reconstitution of the reporter protein activity (Figure 2A). Here, we use a PCA based on a variant of mouse dihydrofolate reductase (mDHFR, henceforth DHFR) enzyme (Tarassov et al., 2008), which was engineered to be resistant to the DHFR inhibitor methotrexate. In the presence of methotrexate, the essential yeast *sc*DHFR is inhibited so that cells do not grow. If, however, cells express the methotrexate-resistant DHFR PCA fragments, fused to two interacting proteins, then cells will grow. A collection of yeast strains was previously created (Tarassov et al., 2008), in which over 4,500 open reading frames are endogenously tagged with each complementary fragment of DHFR, thereby enabling high-throughput measurement of the interactions between a protein of interest fused to the DHFR N-terminal fragment (F[1,2]) and the rest of the proteome fused to the DHFR C-terminal fragment (F[3]) (Figure 2B). Importantly, the folding of the DHFR fragments was shown to be reversible and is likely not to alter equilibrium dissociation constants or kinetics of protein-protein interactions (Remy and Michnick, 1999; Tarassov et al., 2008). In other words, the free energy of the complementation is small relative to the thermal energy, such that the refolded DHFR does not "trap" the proteins probed (Figure S1). Furthermore, refolding of DHFR from fragments is sterically limited by the length of polypeptide linkers between the fragments and proteins that bring them together; in the case of the DHFR, to distances of 8 nm between the tagged termini (Remy and Michnick, 1999; Tarassov et al., 2008). This assures that measurements of local interactions are of high spatial resolution.

To assess the quantitative nature of the DHFR PCA, we measured whether the growth of a strain expressing two proteins X and Y (each fused to one of the two complementary fragments of DHFR) was directly proportional to the concentration of complexes between X and Y, or [X•Y]. This exercise is difficult, because [X•Y] depends on the concentration of X that is accessible to Y and vice versa and on the affinity of X for Y. To simplify the task, we minimized the importance of affinity by using the strong glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter to drive the expression of a neutral protein reporter, the *Venus* (Nagai et al., 2002) yellow fluorescent protein (YFP) fused to F[1,2]. The cassette was carried on a single-copy plasmid in a *MAT***a** strain BY4741 (Experimental Procedures). This strain was mated to *MAT*α strain BY4742 harboring the F[3] fragment coding sequence fused to the open reading frames encoding 83% of the *S. cerevisiae* genes (Tarassov et al., 2008).

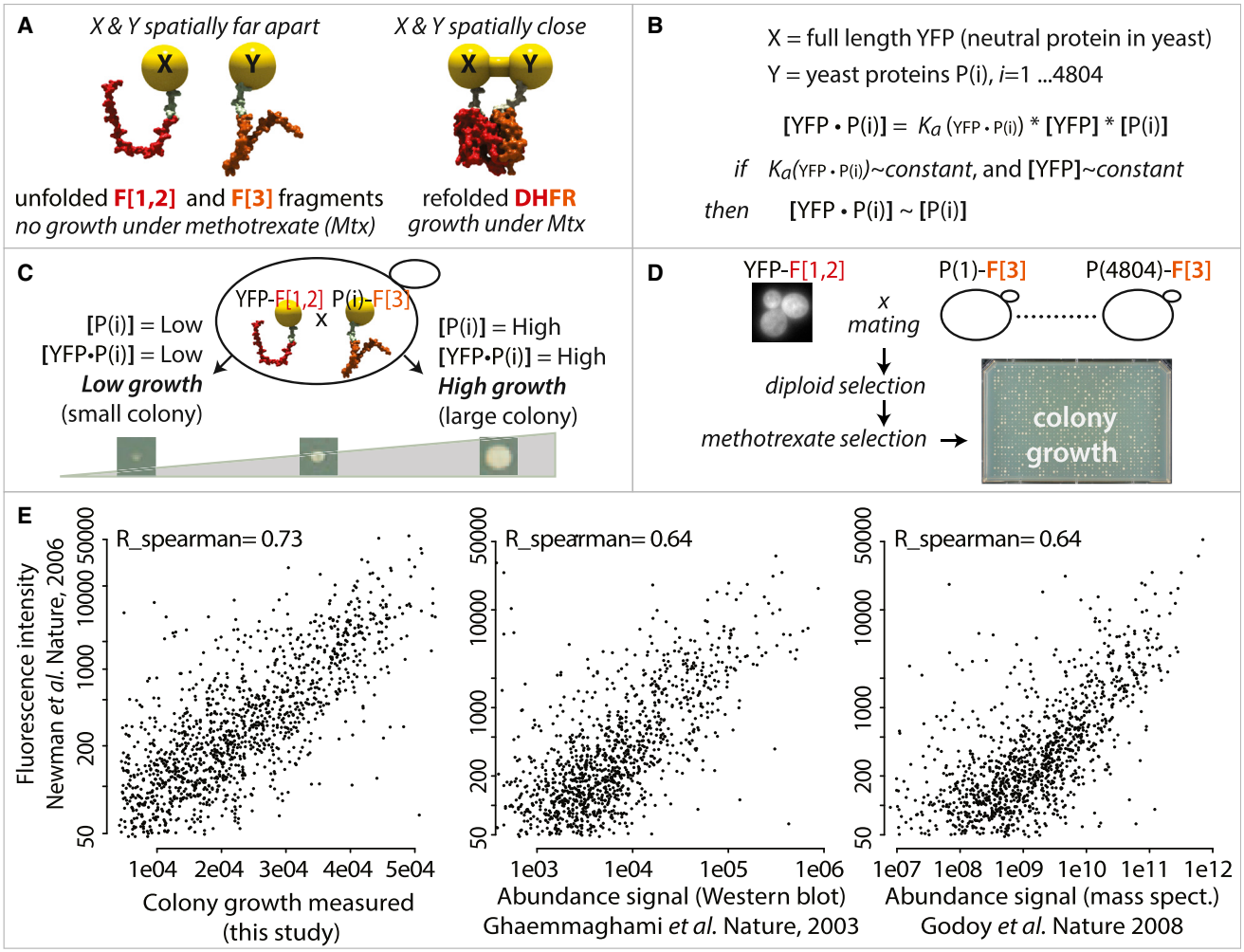

Figure 2. Measuring Protein Concentrations through the Strength of Nonspecific Interactions of a Protein Agent

(A) Principle of the DHFR PCA. Two complementary N- and C-terminal fragments of DHFR (denoted F[1,2] and F[3]) are fused to two proteins of interest, X and Y. Depending on the spatial distance between X and Y in the cell, the fragments may or may not fold together and reconstitute DHFR activity that is necessary for cell growth under methotrexate selection.

(B) We used the *Venus* variant yellow fluorescent protein (YFP) as a neutral "agent" protein and measured its interactions with the yeast proteome, denoted P(i). The concentration of complexes [YFP • P(i)] depends on three parameters: [YFP], [P(i)], and the affinity between them: K_a(YFP • P(i)). YFP is not expected to exhibit specific interactions with P(i), so we assume K_a(YFP • P(i)) to be comparable across all P(i); we discuss this assumption in Supplemental Experimental Procedures (text 1). Moreover, because YFP is constitutively expressed from the yeast genome, we assume [YFP] to be comparable across all diploid strains. Under these assumptions, we predict that [YFP • P(i)] should be proportional to [P(i)].

(C) We hypothesize that the assay is quantitative, or that strain growth is proportional to the concentration of complexes being measured through DHFR reconstitution.

(D) We measure growth on methotrexate for 4,804 diploid yeast strains expressing YFP-F[1,2] and P(i)-F[3].

(E) The left panel shows strain growth versus the abundance of the corresponding protein as measured by fluorescence (Newman et al., 2006). Center and right panels show the abundance measured by western blot (Ghaemmaghami et al., 2003) and mass spectrometry (de Godoy et al., 2008), respectively, versus the abundance measured by fluorescence (Newman et al., 2006), for the same subset of proteins. All versus all methods comparison is shown in Figure S2. The abundance data shown in this graph is available for download as Data Set S2.

The resulting diploid strains were then used to screen for protein-protein interactions following methotrexate selection (Figures 2C and 2D).

Because the sole role of the YFP-F[1,2] chimeric protein is to be a spy from within the cell, reporting on a protein's whereabouts, we termed it the "agent." We reasoned that the fluorescent protein is foreign to yeast and therefore should exhibit only weak and nonspecific interactions with the rest of the yeast proteome. If we assume that affinity constants between the agent and the rest of the yeast proteome are low and comparable to each other as discussed in Supplemental Experimental Procedures (text 1), then the concentration of complexes between the agent and the proteins probed should depend only on their local concentration (Figure 2B). In summary, we would expect

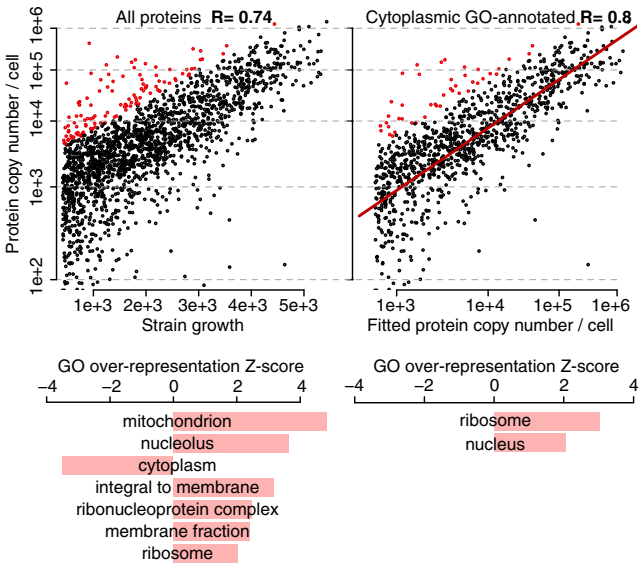

Figure 3. The Relationship between Protein Abundance and Colony Growth Depends on Subcellular Localization

The left panel shows strain growth versus protein abundance from the PAX-db compendium, which we rescaled to reflect copy numbers per cell (see Experimental Procedures). We focus on proteins for which the abundance is underestimated by colony growth (red dots). For these, GO terms are enriched in annotations reflecting subcellular localization and are depleted in the term "cytoplasm." This led us to plot the same data in the right panel, but only considering proteins with a GO annotation containing the term "cytoplasm." This exercise improved the correlation to R = 0.8. After a linear fit of these data (red line), we transformed colony growth into protein abundance to ease the interpretation of quantities in the rest of the manuscript. Many proteins have several GO annotations and those sharing "ribosome" and "nucleus" on top of "cytoplasm" are enriched among proteins for which abundance is underestimated by our method.

that the growth of a yeast strain would depend solely on the abundance of its tagged protein (Figure 2C).

After optimizing the assay to achieve the maximum dynamic range of differences in growth under methotrexate selection (Figure S2), we were able to measure significant growth for ~2,000 strains. This means that a protein constitutively and ubiquitously expressed in yeast has the potential to interact with at least 2,000 other proteins. Although such a large number of interactions can be puzzling, we note that it may be expected if the interactions detected are weak enough. These interactions are indeed fundamentally different from those detected by methods such as affinity purification (Gavin et al., 2006; Krogan et al., 2006) or yeast-two-hybrid (Yu et al., 2008), where the aim is to identify high-affinity, specific interactions. In contrast, our goal is to measure nonspecific protein encounters and, in fact, the number of interactions that we observed (~2,000) is probably underestimated due to, e.g., inevitable mislocalization, misfolding, or degradation of some proteins caused by the F[3] tag fusion. Moreover, ~1,500 genes were not tagged in the library. Yet, despite its likely underestimation, this number indicates that a large fraction of the proteome is physically accessible to the agent.

Given the accessibility of proteins to the agent, we asked whether the frequency at which the agent would interact with a given protein could simply depend on the abundance of that protein. We thus plotted the observed strain growth against the known protein abundances (Newman et al., 2006), which revealed a striking correlation (R = 0.73; Figure 2E; we used only the Spearman correlation coefficient, because it does not depend on the scaling of the data). Because known abundances depend on the method being used to measure them, we compared the agreement between the values observed with our strategy and values measured by other strategies: fluorescence (Newman et al., 2006), western blot (Ghaemmaghami et al., 2003), and mass spectrometry (de Godoy et al., 2008; Figure S2). Remarkably, the correlation that we report above (0.73) is higher than the correlation observed between any other pair of methods (Figure S3). Such consistency observed on a proteome-wide scale can only be explained if the DHFR PCA supports growth in proportion to the concentration of the complex formed, therefore highlighting the quantitative nature of the assay. These results also indicate that the frequency at which proteins encounter each other in a cell primarily depends on their abundance.

The Cytosolic Agent Best Detects Cytosolic Proteins

The availability of proteins in the cytosol is expected to scale with their abundance, but the availability of proteins in other compartments should not obey this relationship. Consistent with this assumption, abundance was underestimated mostly for proteins annotated to subcellular compartments (Figure 3). Enrichment for cytosolic proteins (annotated as cytoplasmic in the Gene Ontology [GO]) improved the correlation between strain growth and abundance (R = 0.8; Figure 3). Furthermore, a GO analysis showed that among proteins annotated as cytoplasmic, those whose abundance is most underestimated by our assay were also enriched in annotations indicative of a subcellular structure (nucleus and ribosome). It is thus possible that a subpopulation of these proteins resides in the nucleus, thereby lowering the available cytosolic concentration.

Partitioning the Cell into Two Spaces

The fact that localized proteins have their abundance underestimated by the cytosolic agent led us to test whether localizing the agent would result in overestimating protein abundance at that location. We thus added a sequence coding for the PMP2 transmembrane helix to the N terminus of YFP-F[1,2] to drive the agent to membranes (see Experimental Procedures). Considering the binding equilibrium formulae (Figure 2B), the concentration of complexes between the agent and any protein is directly dependent on both of their concentrations. Targeting the agent to a particular location should therefore result in an increase in signal with proteins present at that same location and in a decrease in signal with proteins located elsewhere. In other words, the agents report on local protein concentration, such that we expect the membrane agent to report a higher effective abundance than the cytosolic agent for membrane proteins and vice versa. Using the correspondence between growth and protein abundance established in Figure 3, we inferred the local abundance of proteins interacting with the membrane agent and compared it to the local abundance measured by the cytosolic agent.

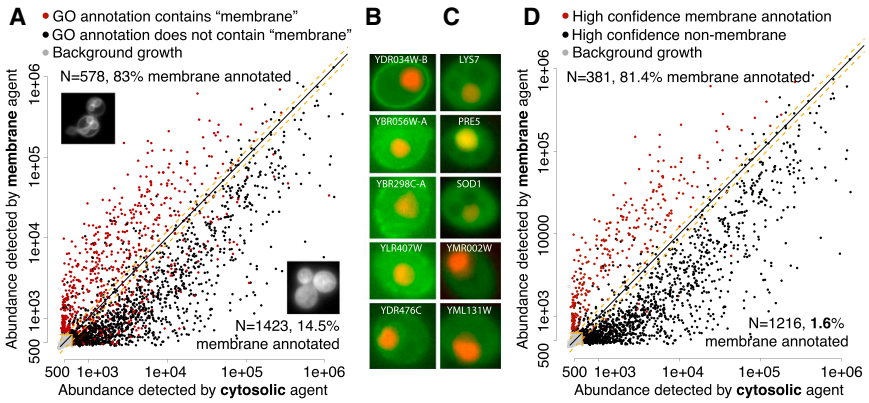

Figure 4. Impact of a Change in Localization on the Encounter Frequencies of a Protein with a Proteome

(A) Abundance measured by the cytosolic agent (x axis) versus abundance measured by the membrane agent (y axis). Of the 578 proteins for which abundance is overestimated by the membrane agent relative to the cytosolic one, 83% contain the term "membrane" in their GO annotation. Conversely, of the 1,423 proteins yielding a stronger signal with the cytosolic agent, only 14.5% contain "membrane" in their GO annotation.

(B) Proteins that showed increases in DHFR PCA signal with the membrane-directed over the cytoplasmic agent were tagged with *Venus* YFP. The majority of these indeed show membrane localization.

(C) Membrane-annotated proteins yielding a stronger signal with the cytosolic agent were tagged and observed. The nucleus is highlighted by a red fluorescent protein-tagged histone protein. Images of all strains are available in Table S1.

(D) Same as (A) using high-confidence data sets for membrane and nonmembrane proteins.

There were 578 proteins for which the membrane agent measured an apparent abundance at least 1.2-fold higher than the cytosolic agent. Of these, 83% exhibited a GO annotation containing the keyword "membrane." Of the remaining 17%, we fluorescently tagged 34 proteins exhibiting a greater than 2-fold change (see Experimental Procedures). These data, together with literature curation, showed that 25 of these 34 proteins are likely to reside at a membrane or be associated with a membrane-rich organelle such as the endoplasmic reticulum (Table S1). Conversely, 1,423 proteins exhibited a higher apparent abundance with the cytosolic agent. Of these, only 14.5% contained "membrane" in their GO annotations. Similarly, tagging 15 of these proteins showed that 10 of them appeared mostly in the cytosol or in the nucleus (Table S2). We show examples where the localization inferred by the agent is confirmed by fluorescence imaging (Figure 4B and 4C), and we show in Tables S1 and S2 all the strains tested. Furthermore, when using a more strictly defined data set of membrane and nonmembrane proteins based on a recent study (Babu et al., 2012), the percentage of membrane-related proteins detected more strongly by the cytosolic agent dropped from 14.5% to 1.6%, suggesting that the agents provide a quantitative description of protein localization. Overall, these findings reflect that the two agents measure the effective concentration of proteins present in their local environment, which enabled us to partition the cell into two spaces. This result is promising for future developments, because the number and the nature of the environments probed depend only on the construction of agents targeted to diverse compartments. Interestingly, the in vivo nature of this strategy would also be compatible to probe the environment of different liquid phases that proteins can form (Li et al., 2012).

Parameters Influencing Protein Encounter Frequencies in Cells

Heterogeneities in protein concentration measured by the two agents are particularly interesting to consider in the context of cellular adaptation, which typically involves changes in protein abundance and localization (Breker et al., 2013; Dénervaud et al., 2013; Ideker and Krogan, 2012; Tkach et al., 2012). We thus sought to compare the contribution that protein abundance and protein localization might have on restructuring the yeast proteome during adaptation. The impact of changes in protein abundance can be simply summarized by the distribution of changes observed across different environmental conditions. For instance, in yeast cells cultured in minimal versus complete media (Newman et al., 2006), the magnitude of changes in protein abundance were relatively small, with only 29 proteins out of 2,196 measured exhibiting an absolute fold-change above 4 (Figure 5A). In another experiment comparing the proteome of haploid versus diploid yeast cells (de Godoy et al., 2008), an even smaller fraction of proteins was observed to undergo such a change (31 out of 3,995) (Figure 5A).

A change in absolute abundance globally increases or decreases protein concentration. In contrast, a change in localization results in a redistribution of local concentrations, with an increase and a decrease at the target and source location, respectively. To compare the magnitude of this redistribution to typical changes in global concentration, we plot the log ratios of the abundances perceived by the membrane and cytosolic agents. For example, the cytosolic agent detects the proteasome subunit Pre9 with an abundance equivalent to ~67,000 copies per cell, whereas the membrane agent detects Pre9 with an abundance equivalent to ~7,000 copies per cell. Therefore, if a given protein were relocalized from the cytosol to the membrane, then it would perceive Pre9 as if it had been downregulated ~9.5-fold. The distribution of these changes across all proteins (Figure 5A) reveals that changes in localization may have a greater impact on the cellular architecture than changes in abundance. We indeed observed that 566 proteins are detected with more than a 4-fold difference between the two agents. The frequency at which such relocalization events occur during adaptation was recently estimated (Breker et al., 2013). Interestingly, more proteins underwent a change in localization than a change in abundance over 2-fold. Together, these results suggest that relocalization significantly contributes to the reorganization of a proteome, perhaps more than changes in abundance alone (Figure 5B).

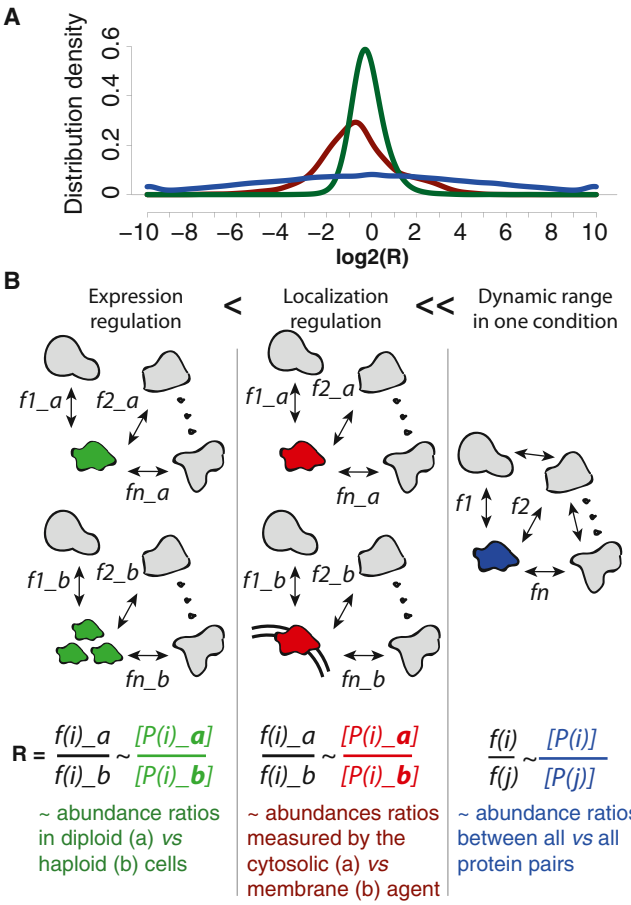

Figure 5. Perception of the Cellular Milieu by Proteins at Steady State upon Adaptation

(A) The cytosolic and membrane agents enable us to measure two environments a protein would perceive should it be cytosolic or at the membrane. We quantify the magnitude of differences between these two environments by the distribution of log ratios (R) of abundances measured by each agent (red). This ratio reflects changes in encounter frequencies with the proteome that a protein would undergo when changing its localization between the cytosol and membrane. For comparison, we plot the density distribution of ratios between absolute abundances measured under two conditions (green) (Newman et al., 2006). Finally, we plot the density distribution of ratios between absolute abundances of all protein pairs in the yeast proteome (blue).

(B) Schematic description of the result shown in (A). Each scenario depicts a given protein (green, red, or blue) that can encounter other proteins P(i), i = 1,2 … n, with frequencies f(i), that are proportional to the concentrations [P(i)]. Encounter frequencies between a protein P(i) and the rest of a proteome can change due to P(i) changing in concentration (left panel, condition a or b) or localization (middle panel, localization a or b). Yet it is remarkable that the distribution of encounter frequencies at equilibrium between protein pairs (approximated by the ratio between values randomly sampled from [P(i)]) is much wider than the distribution of changes in these encounter frequencies observed when expression or localization of proteins change (blue curve in A and right panel, B). This highlights that absolute protein abundance should have an important weight in driving the evolution of cellular systems, because it appears as the main determinant of "who frequently encounters whom." However, for the modulation or regulation of encounter frequencies, we see that changes in localization have a larger impact than changes in expression.

At the same time, when considering a cell population at equilibrium, it is remarkable that the heterogeneity introduced by localization can be counterbalanced by a large abundance. This is also illustrated by the protein Pre9: although it is detected as less abundant by the membrane agent, the abundance detected is still greater than that of many membrane proteins. In other words, because protein abundance spans three to five orders of magnitude, the "leakage" of only 1% of the population out of a particular localization (L1) to another localization (L2) may yield an abundance that is greater than that of many proteins specifically localized in L2. More generally, this example illustrates that the efficacy of protein segregation does not fully compensate for the wide dynamic range of protein abundances. We observe this effect when considering the cytosol-membrane dichotomy, and it will need to be investigated further for other compartments.

DISCUSSION

Our results highlight the cellular architecture as being different from the construction of designed and engineered systems or devices, in which most components carry out their functional or structural roles without physically encountering the majority of the other components. The cell instead appears akin to our society, in which organization is fuzzy and best viewed as a continuum; i.e., most people live in a specific place but can be seen to be in other places from time to time and can potentially be seen anywhere with a given probability. Such a view has at least two important implications, one practical and one conceptual. Practically, it means that quantitative data (concentration, localization, affinity) is required to rationalize protein networks. In an engineered system, for example, knowing that component X interacts with component Y is often sufficient to link X and Y both spatially and functionally. However, in cellular systems, knowing that protein X interacts with protein Y may have different degrees of functional implications depending on the specific properties of the interaction, of the proteins, and of their context. It will thus be important to quantify the interactions between proteins in terms of their strength as well as when and where they take place in the cell. Conceptually, the high intervisibility among proteins provides a large pool of interactions that evolution may adjust through mutation-selection cycles to ultimately yield new functional protein-protein interactions (Kuriyan and Eisenberg, 2007).

EXPERIMENTAL PROCEDURES

PCA Constructs

The two constructs (GPDprom-Venus-*linker*-DH and GPDprom-PMP2-*linker*-Venus-*linker*-DH) are derived from the p413 vector (Mumberg et al., 1995). The sequences of the final constructs are given in Supplemental Experimental Procedures (text 2). Both plasmids were transformed in a BY4741 strain (*MAT***a** *his3*Δ *leu2*Δ *met15*Δ *ura3*Δ) using the same protocol as in Tarassov et al. (2008).

Fluorescent Constructs

Plasmids used to tag genes with Venus (in BY4741) and mCherry (in BY4742) were derived from the plasmids carrying the final homologous recombination cassettes referred to as F[1,2]-NAT1 and F[3]-HPH in Tarassov et al. (2008), where the DHFR fragments were replaced by the fluorescent protein. The sequence of these cassettes is given in Supplemental Experimental Procedures (text 2). Note that some of the strains created do exist in the GFP

collection (Huh et al., 2003), but we reasoned that the *Venus* fluorescent protein would provide higher signal as well as a cross-validation of the tagging process.

DHFR PCA

For all steps, growth is achieved at 30 degrees. Individual strains (BY4741 harboring GPDprom-Venus-F[1,2] and GPDprom-PMP2-Venus- F[1,2]) were grown in liquid (synthetic defined media, −his) for 48 hr and mated onto YPD plates with the collection of yeast strains BY4742 (*MATα his3Δ leu2Δ lys2Δ ura3Δ*) endogenously tagged with the DHFR C-terminal fragment (F[3]). The mating and all subsequent steps were carried out in 1,536 colony format and replication was achieved by a robotic arm operating a 1,536-format pintool (V&P Scientific, FP1N pins). Mated colonies were transferred onto diploid selection plates (synthetic defined media, −his, −met, −lys, +hygromycin), allowed to grow for 2 days, and retransferred a second time onto identical new plates to obtain homogeneous diploid colonies in terms of shape and size. After 1 day of growth, colonies were printed in four replicates onto selection plates containing methotrexate (Mtx) (synthetic defined media, −Ade, +Mtx). Plates were photographed after 3 days of growth with a digital camera. Colonies intensities were measured from the digital pictures using ImageJ "integrated intensity" after 8-bit conversion and after a background correction was applied onto the entire plate (ball radius 50 px). Colony intensities across the four replicates were averaged to yield the "colony growth" used in the main text. A colony was identified as correctly printed when the colony intensity was above 500. We therefore did not consider for analysis those proteins with average colony intensity below 500 in any of the experiments. The range 500–4,000 was considered background growth, and a signal over 4,000 was considered above background.

Microscopy

Strains were inoculated from glycerol stock in a 96-well format in 80 μl of synthetic defined media (−Ade). When optical density at 600 nm reached 0.5–1.5, 4 μl was transferred into a 96-well glass-bottom plate containing 50 μl of the same medium for imaging. Imaging was carried out using a Nikon TE2000E microscope, 60× oil-immersion objective (numerical aperture 1.4, plan apo, Nikon).

Bioinformatics Analyses

Colony intensities were loaded in the R statistical environment for analysis (Ihaka and Gentleman, 1996). GO annotations were retrieved from the *Saccharomyces* Genome Database database as of April 2012 (Cherry et al., 2012). Protein abundance information was obtained from the Pax-db database (Wang et al., 2012), which provides both abundance of individual studies as well as a consensus abundance. In order to relate the Pax-db abundance to protein copy numbers per cell, we used the copy numbers estimated in Ghaemmaghami et al. (2003). Linear regression between the data from Pax-db and data from Ghaemmaghami et al. (2003) results in the following relationship: log10(Pax) = −1.848 × log10(Ghaemmaghami). We thus multiplied all values from Pax-db by $10^{1.848} = 70.46$.

The GO annotations used are provided in Data Set S1. Cytosolic proteins used in Figure 3 were retrieved by taking all proteins with evidence for the "cytoplasm" GO annotation. Membrane proteins used in Figure 4A correspond to those for which any experimental or predicted evidence contains the string "membran" (without the final letter "e"). For Figure 4D, the high-confidence data set of proteins belonging or not to the membrane was derived from Table S2 in Babu et al. (2012). Nonmembrane proteins were those with no predicted transmembrane helix or membrane-related sequence (columns 32–38 = 0) and with negative annotation at the endoplasmic reticulum, endosomes, Golgi, lipid granules, membrane, secretory vesicles, or vacuole (columns 16–20; 25–27 = 0).

The GO term enrichment calculated for Figure 3 was obtained by drawing 1,000 random sets of proteins (with the same number of proteins as in the set being assessed) and counting the number of times (N) each term was observed. The Z score of a particular term was obtained by Z = (Nobs − Nexp)/SD, where SD is the standard deviation of the number of times the term is observed across the 1,000 repeats, Nexp is the mean, and Nobs is the counts in the set being assessed.

Distributions in Figure 5A were obtained from either the log2 ratio of protein abundances measured in YPD and yeast minimal dextrose (green) or using the abundance ratios measured by the membrane or cytosolic agents (red) or using random protein pairs and taking the ratio of fitted-abundances from Pax-db (blue).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, two tables, and three data sets and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.04.009.

AUTHOR CONTRIBUTIONS

E.D.L. and S.W.M. designed the research. J.K. tagged the strains used for microscopy analyses with help from E.D.L. and performed DHFR PCA with help from E.D.L. E.D.L. performed all other experiments and analyzed the data. E.D.L. and S.W.M. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Blanche Schwappach for suggesting the use of the PMP2 helix as a membrane targeting sequence. We thank Debbie Fass, Maya Schuldiner, Sarah Teichmann, Amnon Horovitz, and Ron Milo for their comments on the manuscript. The authors acknowledge support from CIHR grants MOP-GMX-152556 and MOP-GMX- 231013 and NSERC of Canada grant 194582 (to S.W.M.). E.D.L. acknowledges the Human Frontier Science Program for a long-term postdoctoral fellowship as well as the Weizmann Institute of Science for financial support.

Received: August 2, 2013
Revised: February 11, 2014
Accepted: April 7, 2014
Published: May 8, 2014

REFERENCES

Babu, M., Vlasblom, J., Pu, S., Guo, X., Graham, C., Bean, B.D., Burston, H.E., Vizeacoumar, F.J., Snider, J., Phanse, S., et al. (2012). Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. Nature *489*, 585–589.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043–1046.

Breker, M., Gymrek, M., and Schuldiner, M. (2013). A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. J. Cell Biol. *200*, 839–850.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40* (Database issue), D700–D705.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S., and Krogan, N.J. (2007). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol. Cell. Proteomics *6*, 439–450.

de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Fröhlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature *455*, 1251–1254.

Dénervaud, N., Becker, J., Delgado-Gonzalo, R., Damay, P., Rajkumar, A.S., Unser, M., Shore, D., Naef, F., and Maerkl, S.J. (2013). A chemostat array enables the spatio-temporal analysis of the yeast proteome. Proc. Natl. Acad. Sci. USA *110*, 15842–15847.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., and Garrels, J.I. (1999). A sampling of the yeast proteome. Mol. Cell. Biol. *19*, 7357–7368.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241–4257.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Ge, H., Liu, Z., Church, G.M., and Vidal, M. (2001). Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nat. Genet. *29*, 482–486.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., and Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88–93.

Huberts, D.H., and van der Klei, I.J. (2010). Moonlighting proteins: an intriguing mode of multitasking. Biochim. Biophys. Acta *1803*, 520–525.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Ideker, T., and Krogan, N.J. (2012). Differential network biology. Mol. Syst. Biol. *8*, 565.

Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. J. Comput. Graph. Statist. *5*, 299–314.

Janin, J., Bahadur, R.P., and Chakrabarti, P. (2008). Protein-protein interaction and quaternary structure. Q. Rev. Biophys. *41*, 133–180.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., et al. (2002). Subcellular localization of the yeast proteome. Genes Dev. *16*, 707–719.

Kuriyan, J., and Eisenberg, D. (2007). The origin of protein interactions and allostery in colocalization. Nature *450*, 983–990.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

McGuffee, S.R., and Elcock, A.H. (2010). Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput. Biol. *6*, e1000694.

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. Bioessays *35*, 1050–1055.

Mumberg, D., Müller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119–122.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat. Biotechnol. *20*, 87–90.

Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. *37*, 825–831.

Remy, I., and Michnick, S.W. (1999). Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. Proc. Natl. Acad. Sci. USA *96*, 5394–5399.

Rives, A.W., and Galitski, T. (2003). Modular organization of cellular networks. Proc. Natl. Acad. Sci. USA *100*, 1128–1133.

Simonis, N., Gonze, D., Orsi, C., van Helden, J., and Wodak, S.J. (2006). Modularity of the transcriptional response of protein complexes in yeast. J. Mol. Biol. *363*, 589–610.

Spirin, V., and Mirny, L.A. (2003). Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci. USA *100*, 12123–12128.

Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X., Shi, X., et al. (2011). The BioGRID Interaction Database: 2011 update. Nucleic Acids Res. *39* (Database issue), D698–D704.

Tan, K., Shlomi, T., Feizi, H., Ideker, T., and Sharan, R. (2007). Transcriptional regulation of protein complexes within and across species. Proc. Natl. Acad. Sci. USA *104*, 1283–1288.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science *320*, 1465–1470.

Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C., et al. (2012). Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. Nat. Cell Biol. *14*, 966–976.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417*, 399–403.

Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S.P., Hengartner, M.O., and von Mering, C. (2012). PaxDb, a database of protein abundance averages across all three domains of life. Mol. Cell. Proteomics *11*, 492–500.

Warner, J.R., and McIntosh, K.B. (2009). How common are extraribosomal functions of ribosomal proteins? Mol. Cell *34*, 3–11.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104–110.