9. Gatz, C. Chemical control of gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48,** 89–108 (1997).
10. Gatz, C., Frohberg, C. & Wendenburg, R. Stringent repression and homogeneous de-repression by tetracycline of a modified CaMV 35S promoter in intact transgenic tobacco plants. *Plant J.* **2,** 397–404 (1992).
11. Raschke, K. Stomatal action. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **26,** 309–340 (1975).
12. Mansfield, T. A., Hetherington, A. M. & Atkinson, C. J. Some current aspects of stomatal physiology. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41,** 55–75 (1990).
13. Hedrich, R. *et al.* Changes in apoplastic pH and membrane potential in leaves in relation to stomatal responses to CO_2, malate, abscisic acid or interruption of water supply. *Planta* **213,** 594–601 (2001).
14. Hedrich, R. *et al.* Malate-sensitive anion channels enable guard-cells to sense changes in the ambient CO_2 concentration. *Plant J.* **6,** 741–748 (1994).
15. Otto, B. & Kaldenhoff, R. Cell-specific expression of the mercury-insensitive plasma-membrane aquaporin NtAQP1 from *Nicotiana tabacum. Planta* **211,** 167–172 (2000).
16. Raschke, K. Saturation kinetics of velocity of stomatal closing in response to CO_2. *Plant Physiol.* **49,** 229–234 (1972).
17. Gallois, P. & Marinho, P. Leaf disk transformation using *Agrobacterium tumefaciens*—expression of heterologous genes in tobacco. *Methods Mol. Biol.* **49,** 39–48 (1995).
18. Kaiser, W. M. Correlation between changes in photosynthetic activity and changes in total protoplast volume in leaf tissue from hygro-, meso- and xerophytes under osmotic stress. *Planta* **154,** 538–545 (1982).

Acknowledgements We thank W. M. Kaiser, M. Eckert and A. Schubert for discussion and help in experimental design.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (kaldenhoff@bio.tu-darmstadt.de).

Global analysis of protein expression in yeast

Sina Ghaemmaghami[1,2], **Won-Ki Huh**[1,3], **Kiowa Bower**[1,2], **Russell W. Howson**[1,3], **Archana Belle**[1,3], **Noah Dephoure**[1,3], **Erin K. O'Shea**[1,3] **& Jonathan S. Weissman**[1,2]

[1] *Howard Hughes Medical Institute,* [2] *Departments of Cellular & Molecular Pharmacology and* [3] *Biochemistry & Biophysics, University of California–San Francisco, San Francisco, California 94143-2240, USA*

The availability of complete genomic sequences and technologies that allow comprehensive analysis of global expression profiles of messenger RNA[1–3] have greatly expanded our ability to monitor the internal state of a cell. Yet biological systems ultimately need to be explained in terms of the activity, regulation and modification of proteins—and the ubiquitous occurrence of post-transcriptional regulation makes mRNA an imperfect proxy for such information. To facilitate global protein analyses, we have created a *Saccharomyces cerevisiae* fusion library where each open reading frame is tagged with a high-affinity epitope and expressed from its natural chromosomal location. Through immunodetection of the common tag, we obtain a census of proteins expressed during log-phase growth and measurements of their absolute levels. We find that about 80% of the proteome is expressed during normal growth conditions, and, using additional sequence information, we systematically identify mis-annotated genes. The abundance of proteins ranges from fewer than 50 to more than 10^6 molecules per cell. Many of these molecules, including essential proteins and most transcription factors, are present at levels that are not readily detectable by other proteomic techniques nor predictable by mRNA levels or codon bias measurements.

The diverse chemical nature of proteins makes the development of globally applicable proteomic assays very challenging. We have overcome this obstacle in the yeast *S. cerevisiae* by individually tagging each of its annotated open reading frames (ORFs) with a high-affinity epitope tag so that the resulting fusion proteins are expressed under the control of their natural promoters. The fusion library allows the immunodetection and immunopurification of the entire yeast proteome using a single antibody, enabling the development of a range of high-throughput functional assays. To allow for the facile construction of epitope-tagged yeast fusion libraries, we synthesized 6,234 pairs of ORF-specific oligonucleotide primers. Each of the oligonucleotide pairs have shared 3′ ends that allow for polymerase chain reaction (PCR) amplification of a common insertion cassette, as well as gene-specific 5′ ends that allow for the precise introduction, through homologous recombination, of the amplified insertion cassettes as a perfect in-frame fusion at the carboxy-terminal end of the coding region of each gene[4] (Fig. 1a). The insertion cassettes contained the coding region

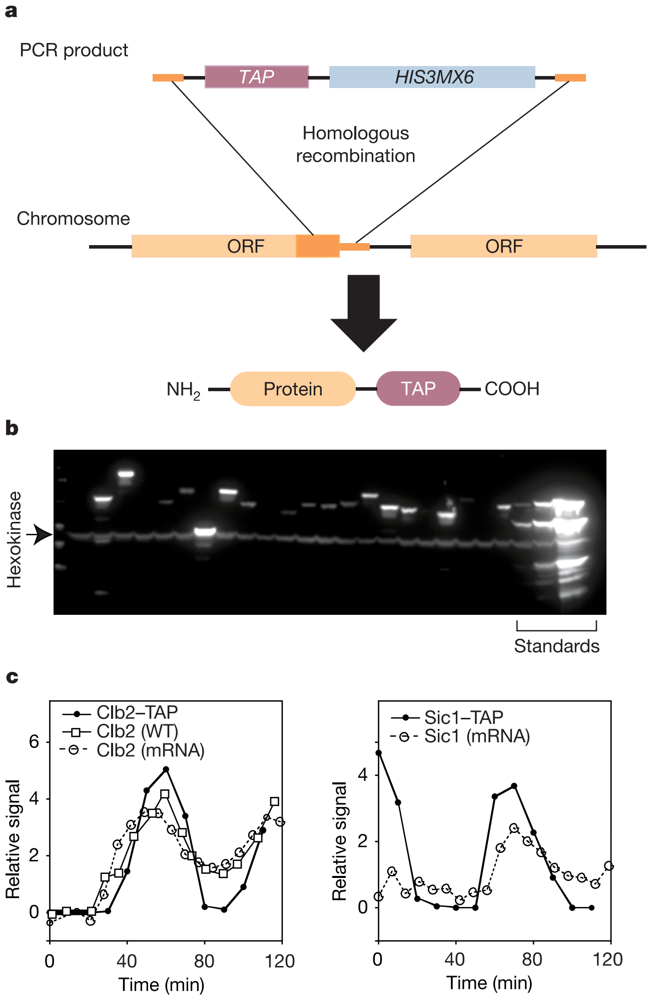

Figure 1 Tagging and detection of the yeast proteome. **a**, Schematic diagram of tagging strategy. **b**, Detection of tagged proteins. Extracts containing TAP-fusion proteins were prepared and analysed by western blots using an anti-CBP antibody (see Supplementary Information). Immunodetection of an endogenous protein (hexokinase) provided a loading control. Serial dilutions of TAP-tagged proteins provided an internal abundance standard (right). **c**, Monitoring dynamic protein levels for two cell-cycle regulated proteins. Strains expressing Clb2– and Sic1–TAP fusions were grown to log-phase and arrested in G1 by α-factor treatment. The cell cycle was resumed by α-factor removal, aliquots were taken at 7-min intervals and levels of the tagged proteins were quantified using western blot analysis (filled circles). For comparison, we include mRNA levels of the two proteins obtained by an earlier microarray analysis[29] (open circles) as well as changes in untagged Clb2 protein levels (open squares) obtained using an antibody against the endogenous protein in an untagged strain.

for a modified version of the tandem affinity purification (TAP) tag[5,6], which consists of a calmodulin binding peptide, a TEV cleavage site and two IgG binding domains of *Staphylococcus aureus* protein A, as well as a selectable marker (see Supplementary Information). In total, we obtained successful integrants for 98% of all ORFs annotated in the Saccharomyces genome database (as of April 2001; http://www-genome.stanford.edu/Saccharomyces), including 93% of all essential ORFs[7] in haploid yeast.

Western blot analysis, using an antibody that specifically recognizes the TAP tag, demonstrated that the large majority (>95%) of detected fusion proteins migrate predominantly as a single band of the approximate expected molecular mass (Fig. 1b). Furthermore, analysis of two known cell-cycle-regulated proteins, Clb2 and Sic1[8,9], indicated that the tagging does not hinder their regulated proteolysis by the ubiquitin/proteasome degradation system and that the TAP tag itself is rapidly destroyed during the targeted degradation of the fusion protein (Fig. 1c). These and other data[6] suggest that the function, regulation and stability of most, but not all (see Supplementary Information), of the proteome is uncompromised by the fused tag.

We observed a protein product for 4,251 of the TAP-tagged ORFs by comprehensive western blot analysis. This set of proteins shows excellent overlap (>90%) with the set of green fluorescent protein (GFP) fusion proteins detected by fluorescence microscopy[10] (Fig. 2a), and together indicate that at least 4,517 proteins are expressed during log-phase growth in rich media. We detect 79% of all essential proteins and 83% of gene products corresponding to ORFs with assigned gene names. By contrast, only 73% of all annotated ORFs expressed a detectable protein product (Fig. 2b). This discrepancy largely results from the presence of spurious ORFs in the annotated yeast genome database stemming from well-known difficulties in distinguishing actual coding regions from fortuitous short ORFs[11,12]. For the original annotation of the yeast genome, an arbitrary cut-off of 100 codons was used to qualify ORFs as potential genes, leading to an anomalous peak centred between 100 and 150 amino acids in the sequence length distribution (Fig. 2c, black)[13] of the genome that is not present in the length distribution of the subset of named genes (Fig. 2c, green). Importantly, although we tagged and analysed all potential ORFs, the length distribution of the subset of observed proteins did not contain the above artefactual peak (Fig. 2c, red), indicating that our analysis of expressed genes has a very low false-positive rate (see also Supplementary Information).

A number of bioinformatics approaches, including recent analyses of the genomic sequences of a number of related yeast species, have been used to distinguish between the real and misannoted ORFs[14–17], although the true number and identity of the spurious ORFs remain unclear. Our results offer experimental verification for a large number of hypothetical genes (we observed 1,018 protein products belonging to functionally uncharacterized ORFs), and yields a large, experimentally validated set to evaluate the success of computational methods for identifying falsely annotated genes. By combining a novel metric—termed the codon enrichment correlation (CEC), which evaluates the patterns of codon usage in potential ORFs—with our protein expression data, we identified a

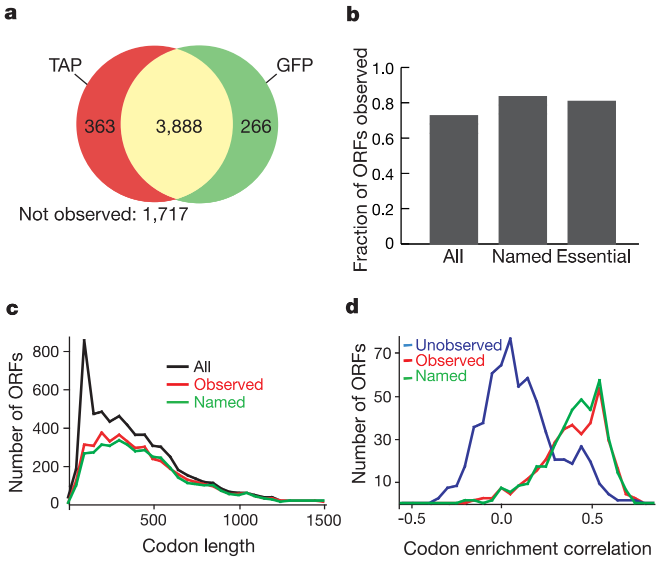

Figure 2 Analysis of proteins expressed during log-phase growth. **a**, Venn diagram comparing sets of proteins detected by western blot of TAP-tagged strains (red), fluorescence microscopy of GFP-tagged strains[10] (green) and both (yellow). **b**, Fraction of the indicated set of ORFs observed in either the TAP-tagged or GFP-tagged libraries. **c**, Size distribution of ORFs, binned by length using 50-codon intervals. The number of ORFs per bin is plotted for the indicated sets of ORFs. **d**, Codon enrichment correlation (CEC) distribution of small ORFs. CECs were calculated for ORFs with lengths from 100 to 150 codons. ORFs were binned according to CEC values using intervals of 0.05 units. The number of ORFs in each bin is plotted for the indicated sets of ORFs. Note, observed proteins have a positive CEC value characteristic of named genes, whereas unobserved ORFs show a major peak centred near a zero value expected for random sequences.

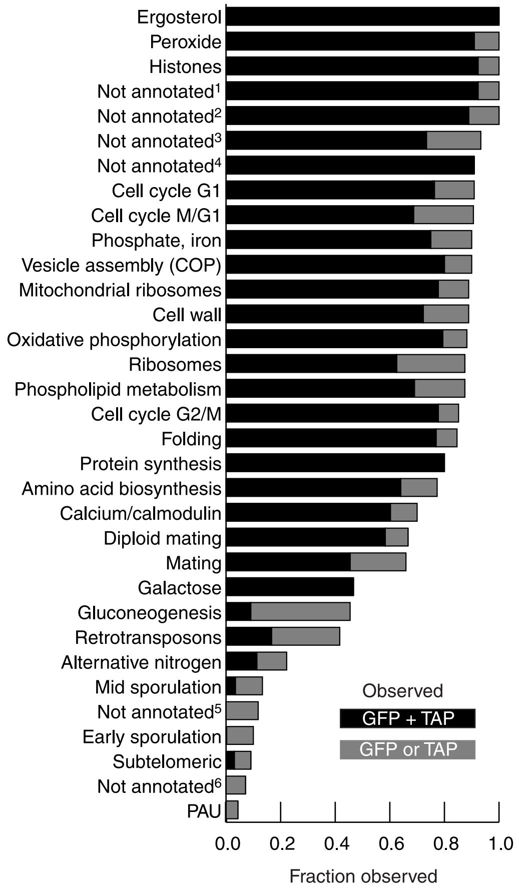

Figure 3 Functional categorization of proteins expressed during log-phase growth in rich medium. 33 modules of co-expressed, functionally related genes were identified by global analysis of ~1,000 microarray data sets[18,19]. Plotted is the fraction of the ORFs in each module that produced a detectable protein product by TAP western analysis or GFP microscopy[10] alone (grey), or both methods (black). Where possible, modules are annotated by function. The gene composition of the modules can be obtained at http://barkai-serv.weizmann.ac.il/modules/ using a cut-off threshold of 4.0. 'Not annotated[1–6]' correspond to modules containing YHR025W, YER103W, YPL016W, YPL180W, YER039C-A and YCL076W, respectively.

set of 525 potentially spurious ORFs (listed in Supplementary Information) that have codon compositions not characteristic of genuine genes and did not yield detectable protein products (Fig. 2d, Methods). On the basis of the CEC distribution of genuine ORFs, we estimate that this list is contaminated by ~20 genuine coding sequences. Our proteomics-based approach complements the comparative genomics strategy for identifying spurious ORFs[16]. The large majority (all but seven) of the 496 spurious ORFs suggested by Kellis *et al.*[16] were not observed in our TAP and GFP studies. The set of spurious ORFs that we identified overlaps well with those detected by this cross-species genome study (381 genes were identified as spurious by both studies), and expands the set by 144 ORFs. Among these 144 ORFs are a large number of sequences that overlap with real genes on the opposite strand, and therefore are difficult to distinguish through homology analysis.

After discounting the spurious ORFs, there remain ~1,000 genuine coding regions that did not produce a detectable protein product. To determine if the unobserved proteins belong to classes of genes that are not transcribed during normal log-phase growth conditions, we compared our results with global transcriptional array data. A recent analysis of mRNA expression profiles from ~1,000 published microarray experiments allowed for the identification of 33 'modules' of transcriptionally co-regulated genes[18,19]. For modules that are expressed in log phase (for example, those coding for housekeeping functions, such as ergosterol and amino-acid biosynthesis and cell cycle), we were able to detect the large majority of the protein products (Fig. 3). By contrast, modules composed of genes involved in functions required only under specialized conditions (for example, meiosis/sporulation and alternative nitrogen utilization) generally produced few detectable proteins.

We took advantage of the fact that all gene products were detected using the same epitope/antibody interaction to measure the absolute abundance of each of the tagged proteins using quantitative western blot analyses. This effort was facilitated by the inclusion of internal standards in each gel (Fig. 1b). We find that the levels of different proteins show an enormous dynamic range, varying from fewer than 50 to more than 10^6 molecules per cell (Fig. 4a, b). The results show that previous efforts to quantify protein levels using two-dimensional gel electrophoresis or mass spectrometry were strongly biased towards the detection of abundant proteins (Fig. 4a, see also Supplementary Fig. S3)[20–23]. For example, a recent study using mass spectrometry and isotope labelling succeeded in quantitatively monitoring changes in the abundance of 688 yeast proteins[22]. For the most abundant proteins (>50,000 molecules per cell) the coverage was excellent (~60%), whereas for the 75% of the proteome that is present at fewer than 5,000 molecules per cell, only 8% of the proteins were observed. Another mass-spectrometry effort that focused on detecting, without directly quantifying, the complement of proteins in log-phase yeast[23] observed a larger number (1,484) of proteins, although it was also biased towards abundant proteins (90% of the proteome present at >50,000

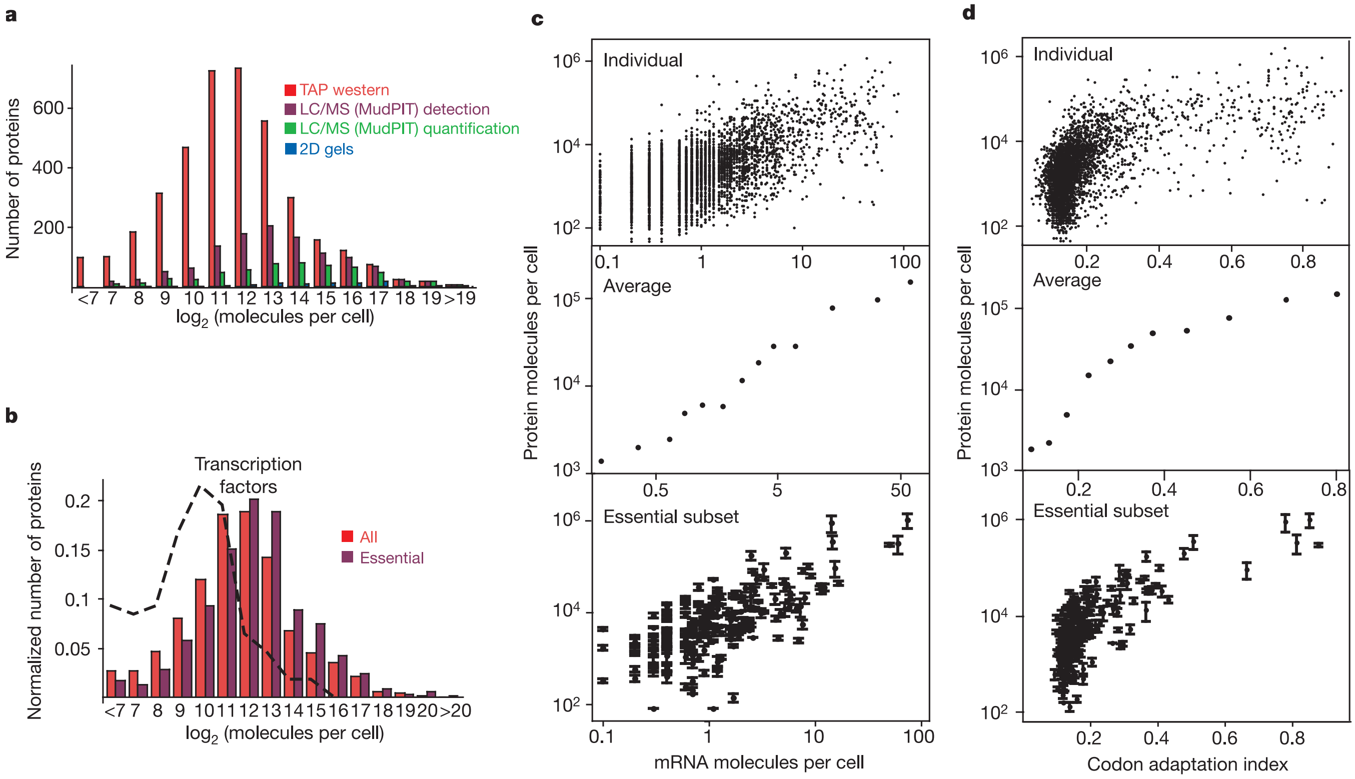

Figure 4 Abundance distribution of the yeast proteome. **a**, Distribution of yeast proteins observed by TAP/western-blot (red), liquid chromatography/mass spectrometry multidimensional protein identification technology (LC/MS MudPIT) analysis focusing on comprehensive detection[23] (purple) and quantitative analysis[22] (green), and combined results from 2 two-dimensional (2D) gel analyses[20,21] (blue). The bins are $\log_2$ increments with upper boundaries indicated. **b**, Normalized abundance distribution of observed proteins (red), essential proteins (purple) and transcription factors (dashed line). **c**, The relationship between steady-state mRNA and protein levels. Top plot, abundance of each protein is plotted against its mRNA level determined by microarray analysis[25]. Middle plot, ORFs are sorted according to mRNA levels, and binned into successive groups with cut-offs of 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 10, 20, 50 and 100 molecules per cell. For each bin, the mean protein abundance is plotted against the mean mRNA level. Bottom plot, protein versus mRNA relationship for a subset of essential soluble proteins (see Supplementary Information). Errors represent the standard deviation of three measurements. **d**, Relationship between codon adaptation index (CAI) and protein levels. Individual and averaged protein values are plotted against CAI[27]. In the middle plot, the values are binned using CAI cut-offs of 0.1, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.60, 0.70, 0.80 and 1.0.

molecules per cell was detected, whereas only 19% of the proteome present at fewer than 5,000 molecules per cell was observed). Our validated list of expressed proteins will help evaluate future advances in mass spectrometry approaches[24].

Overall, we observe a significant relationship between mRNA levels, as measured by an earlier microarray analysis of log-phase yeast[25], and protein levels (Spearman rank correlation coefficient $r_s = 0.57$). Very abundant mRNAs generally encode for abundant proteins, and the average protein per mRNA ratio remains remarkably constant throughout the full range of mRNA abundances (Fig. 4c, middle, and Supplementary Fig. S4). The average protein per mRNA ratio is 4,800 using this measure of mRNA levels, and is 4,200 using an alternative mRNA abundance measurement based on a microarray analysis comparing mRNA to genomic DNA levels[26] (Supplementary Fig. S4). However, individual genes with equivalent mRNA levels can result in large differences in protein abundances (Fig. 4c, top). To assess if this variability was primarily caused by protein measurement error and/or disruption of protein function by the TAP tag, we performed further triplicate measurements of protein abundances on a subset of 206 essential, soluble proteins (See Supplementary Information); the selected strains grew robustly, showing that the tagged proteins were functional. This subset also shows a high degree of protein to mRNA variability relative to our measurement error, indicating that the large differences in individual protein to mRNA ratios are not due primarily to noise in the protein abundance measurements or disruption of the protein by the tag (Fig. 4c, bottom). However, the correlation between mRNA and protein levels is somewhat greater ($r_s = 0.66$), suggesting that the disruption of protein by the TAP tag or difficulty in analysing membrane proteins may have contributed to some of the variation. We also observed a significant relationship ($r_s = 0.55$) between protein abundance and codon usage as measured by the codon adaptation index (CAI)[27]. Protein abundances drop rapidly for genes with CAI values <0.2, explaining the difficulty that previous proteomic approaches have typically had in detecting these proteins[22]. But on an individual gene basis, there is great variability that is also present in the subset of more carefully measured essential, soluble proteins (Fig. 4d).

A number of observations support the argument that the full range of abundances detected in this study, including the very low expression levels, represent functionally significant amounts of the proteins. First, the analysis of transcription modules (Fig. 3) indicates that within groups of genes that are turned off during log-phase growth the corresponding proteins are not observed, even at residual levels. Second, the abundance distribution profile of the entire yeast proteome (Fig. 4b, red) is similar to the profile of the portion of the proteome whose function is required for survival under standard growth conditions (Fig. 4b, purple). This suggests that, in general, functional proteins are not under-represented amongst low-abundance proteins. Third, there are entire classes of functionally important proteins, such as transcription factors (Fig. 4b, line) and cell-cycle proteins (Supplementary Fig. S5), that are present at very low expression levels. Thus the low-abundance proteins detected and quantified in the present study represent a large and functionally important portion of the yeast proteome that is almost entirely invisible to systematic quantitative analysis by other proteomic methods.

The TAP-tagged library now makes it feasible to monitor dynamically the abundance of the yeast proteome through basic cellular events such as the cell cycle and meiosis, and will allow the determination of protein lifetimes. In addition, important subsets of proteins, such as transcription factors, can be readily studied under a more comprehensive set of conditions. This protein-based data will provide critical information for efforts to understand the logic of cellular regulatory circuits, and, by comparison to mRNA levels, the data will give insight into the nature and extent of post-transcriptional regulation. □

Methods

Quantification of protein levels

Cultures (1.7 ml) of tagged strains were grown in 96-well format to log phase, and total cell extracts were examined by SDS–polyacrylamide gel electrophoresis (PAGE)/western blot analysis as described in Supplementary Information. The bands corresponding to the tagged proteins were detected using chemiluminescence and a CCD camera (FluorChem 8800, Alpha Innotech). To control for variation in extraction and loading, each blot was probed with an antibody against endogenous hexokinase in addition to the TAP-specific anti-CBP antibody. Extracts whose hexokinase signals varied by greater than a factor of ~2 from the expected value were re-grown and re-analysed. A standard containing a mixture of three TAP-tagged proteins (Pgk1, Cdc19, Rpl1A) were included in each gel at one-, ten- and 100-fold dilutions. Proteins whose chemiluminescence signals were approaching saturation were re-examined by performing the western blot analysis using a tenfold dilution of the extract and/or lower exposure times during detection. Before the quantitative SDS–PAGE/western blot analysis, strains were ordered on the basis of estimates of TAP abundance from a preliminary dot-blot analysis. In order to provide a standard for the conversion of western signals to absolute protein levels, a TAP-tagged protein (*Escherichia coli* initiation factor A, INFA) was overexpressed in *E. coli* and purified to homogeneity. Yeast extracts containing serial dilutions of INFA ranging from 500 attomoles (which was the limit of detection, see Supplementary Fig. S1) to 25 picomoles were run on a gel along with extracts from 25 different yeast TAP-tagged strains representing the full range of observed protein signals (a second TAP-tagged protein (initiation factor B) was also analysed to ensure that the observed TAP signal was not influenced by the fusion protein). Comparison of the signals generated by these 25 proteins to the known standards allowed the creation of a conversion factor between the observed western blot signals and absolute protein levels. Based on the number of cells ($\sim 1 \times 10^7$) used for the SDS–PAGE/western blot analysis, the protein levels were then converted to measurements of protein molecules per cell.

In order to assess the error in our quantification, a set of 33 proteins with a range of abundances were grown in duplicate cultures, separately extracted and analysed on different gels. The replicate signals showed a linear correlation coefficient of $R = 0.94$, with the pairs of proteins having a median variation of a factor of 2.0. This error analysis does not account for potential alterations in the endogenous levels of the proteins caused by the fused tag, which may be particularly disruptive for small proteins (Supplementary Information) or difficulty in analysing some polytopic membrane proteins by SDS–PAGE. For dynamic measurements of protein levels (for example, the cell-cycle dependence of Clb2 and Sic1 levels shown in Fig. 1c or triplicate measurements in Fig. 4c, d) much smaller errors can be obtained by running the samples being compared side-by-side on a single gel. For quantification in the triplicate measurements shown at the bottom of Fig. 4c, d, serial dilutions of extracts containing purified TAP-tagged INFA were run on each gel.

CEC and identification of spurious ORFs

Codon usage in genuine protein-coding regions deviates systematically from randomly generated ORFs, owing to both preferences in amino-acid composition and biases in the usage of synonymous codons[28], and the codon enrichment correlation (CEC) provides a measure of this deviation. To calculate CEC values, we first determined the relative prevalence of the 61 amino acids specifying codons in the 3,753 named ORFs (Supplementary Table S1). The codon usage expected in random sequences was then calculated based on the approximate prevalence of 30% T, 30% A, 20% C and 20% G nucleotides in the yeast genomes. The enrichment of each codon for the positive set is given by dividing its prevalence among the named ORFs by its expected prevalence in random sequences (Supplementary Table S1). Codon enrichments were similarly calculated for each test ORF. The CEC is the linear correlation coefficient (r) between the codon enrichments of the test ORF and the positive set (for examples, see Supplementary Fig. S2). ORFs were designated as spurious if they failed to be detected by both the TAP and GFP analyses, and they had CEC values below a cut-off of 0.25, 0.16, 0.07 or 0.06 for ORFs of size 0–150, 151–200, 201–250 and 251–300 codons, respectively. For ORFs >150 amino acids, these values were chosen so that $<4.5\%$ of the ORFs falling below these cut-offs that are not detected by the GFP or TAP analyses are genuine coding sequences. The number of genuine coding sequences contaminating our list of spurious ORFs was estimated for each size range and CEC cut-off by the following equation: $N_{\text{real}} = N_{\text{obs}}R$, where N_{obs} is the number of detected ORFs that have a CEC value below the cut-off, and R is the ratio of unobserved to observed ORFs, as determined by the probability of detecting named ORFs for the given size range.

Received 28 July; accepted 28 August 2003; doi:10.1038/nature02046.

1. Lashkari, D. A. *et al.* Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl Acad. Sci. USA* **94,** 13057–13062 (1997).
2. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422,** 835–847 (2003).
3. Lockhart, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14,** 1675–1680 (1996).
4. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14,** 953–961 (1998).
5. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17,** 1030–1032 (1999).
6. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).
7. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).
8. Schwob, E., Bohm, T., Mendenhall, M. D. & Nasmyth, K. The B-type cyclin kinase inhibitor p40SIC1

controls the G1 to S transition in *S. cerevisiae*. *Cell* **79,** 233–244 (1994).

9. Grandin, N. & Reed, S. I. Differential function and expression of *Saccharomyces cerevisiae* B-type cyclins in mitosis and meiosis. *Mol. Cell. Biol.* **13,** 2113–2125 (1993).
10. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425,** 686–691 (2003).
11. Harrison, P. M., Kumar, A., Lang, N., Snyder, M. & Gerstein, M. A question of size: The eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* **30,** 1083–1090 (2002).
12. Goffeau, A. Four years of post-genomic life with 6,000 yeast genes. *FEBS Lett.* **480,** 37–41 (2000).
13. Das, S. *et al.* Biology's new Rosetta stone. *Nature* **385,** 29–30 (1997).
14. Kowalczuk, M., Mackiewicz, P., Gierlik, A., Dudek, M. R. & Cebrat, S. Total number of coding open reading frames in the yeast genome. *Yeast* **15,** 1031–1034 (1999).
15. Zhang, C. T. & Wang, J. Recognition of protein coding genes in the yeast genome at better than 95% accuracy based on the Z curve. *Nucleic Acids Res.* **28,** 2804–2814 (2000).
16. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 (2003).
17. Cliften, P. *et al.* Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science* **301,** 71–76 (2003).
18. Ihmels, J. *et al.* Revealing modular organization in the yeast transcriptional network. *Nature Genet.* **31,** 370–377 (2002).
19. Bergmann, S., Ihmels, J. & Barkai, N. Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E* **67,** 031902 (2003).
20. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19,** 1720–1730 (1999).
21. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. A sampling of the yeast proteome. *Mol. Cell. Biol.* **19,** 7357–7368 (1999).
22. Washburn, M. P. *et al.* Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **100,** 3107–3112 (2003).
23. Washburn, M. P., Wolters, D. & Yates, J. R. III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19,** 242–247 (2001).
24. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
25. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95,** 717–728 (1998).
26. Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl Acad. Sci. USA* **99,** 5860–5865 (2002).
27. Sharp, P. M. & Li, W. H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15,** 1281–1295 (1987).
28. Grantham, R., Gautier, C. & Gouy, M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* **8,** 1893–1912 (1980).
29. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9,** 3273–3297 (1998).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank A. Carroll and F. Sanchez for technical assistance; J. Falvo, L. Gerke, J. Newman and members of the Weissman and O'Shea laboratories for discussions; and N. Barkai for providing data before publication. This work was supported by the Howard Hughes Medical Institute and the David and Lucile Packard Foundation. S.G. is a recipient of the Ruth L. Kirschstein National Research Service Award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.W. (jsw1@itsa.ucsf.edu).

corrigendum

Invariant scaling relations across tree-dominated communities

Brian J. Enquist & Karl J. Niklas

Nature **410**, 655–660 (2001).

Equation (1) of this Article was incorrect as printed. The total biomass, M_{Tot}, per unit area is the summation, or integral, across the size distribution of the number of individuals per unit area, multiplied by their body mass. Thus $M_{\mathrm{Tot}} = \int_a^b MN(M)\mathrm{d}M$. Because the number of individuals in a given area is an allometric function of their size, M, we can substitute the observed relationship $N = C_m M^{-3/4}$ to yield the community biomass equation:

$$M_{\mathrm{Tot}} = \int_a^b C_m M^{1/4} = \frac{4}{5} C_m \left(M_a^{5/4} - M_b^{5/4} \right) \qquad (1)$$

This change does not affect any of the reported conclusions or empirical patterns. □