LETTER

doi:10.1038/nature11835

A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis

Paola Picotti[1,2]*, Mathieu Clément-Ziza[3]*, Henry Lam[4]*, David S. Campbell[5], Alexander Schmidt[6], Eric W. Deutsch[5], Hannes Röst[1], Zhi Sun[5], Oliver Rinner[1,7], Lukas Reiter[1,7], Qin Shen[1,8], Jacob J. Michaelson[3], Andreas Frei[1], Simon Alberti[9], Ulrike Kusebauch[5], Bernd Wollscheid[1], Robert L. Moritz[5], Andreas Beyer[3] & Ruedi Aebersold[1,10,11]

Experience from different fields of life sciences suggests that accessible, complete reference maps of the components of the system under study are highly beneficial research tools. Examples of such maps include libraries of the spectroscopic properties of molecules, or databases of drug structures in analytical or forensic chemistry. Such maps, and methods to navigate them, constitute reliable assays to probe any sample for the presence and amount of molecules contained in the map. So far, attempts to generate such maps for any proteome have failed to reach complete proteome coverage[1–3]. Here we use a strategy based on high-throughput peptide synthesis and mass spectrometry to generate an almost complete reference map (97% of the genome-predicted proteins) of the *Saccharomyces cerevisiae* proteome. We generated two versions of this mass-spectrometric map, one supporting discovery-driven (shotgun)[3,4] and the other supporting hypothesis-driven (targeted)[5,6] proteomic measurements. Together, the two versions of the map constitute a complete set of proteomic assays to support most studies performed with contemporary proteomic technologies. To show the utility of the maps, we applied them to a protein quantitative trait locus (QTL) analysis[7], which requires precise measurement of the same set of peptides over a large number of samples. Protein measurements over 78 *S. cerevisiae* strains revealed a complex relationship between independent genetic loci, influencing the levels of related proteins. Our results suggest that selective pressure favours the acquisition of sets of polymorphisms that adapt protein levels but also maintain the stoichiometry of functionally related pathway members.

In proteomics, the generation of reference maps covering a complete proteome has been attempted in two ways. The first is based on the development of immunoassays to detect target proteins and is exemplified for the human proteome by the Protein Atlas project[8]. The second approach is in-depth mapping of a proteome through the collection of fragment ion spectra from multiple mass-spectrometry-based shotgun proteomic experiments[3,4,9]. Such reference spectra can be used in discovery-driven experiments to analyse acquired fragment ion spectra using spectral matching[10–12], or in targeted measurements, to specifically monitor proteins of interest by selected reaction monitoring (SRM)[5,6,13]. So far, both approaches have failed to reach complete proteome coverage; in the case of yeast[1–3] and other microbes[14,15], saturation is apparent when about two-thirds of all proteins that would be predicted based on the genome are covered, and this coverage is much lower for other proteomes, including the human proteome[1].

We defined the yeast proteome as the collection of 6,607 protein sequences predicted based on the genome of yeast, each one associated with an open reading frame (ORF) in the *Saccharomyces* Genome Database (SGD, http://www.yeastgenome.org). First, we classified yeast proteins based on their detectability, using a large repository of proteomic data, PeptideAtlas[1,3], and the largest data set of antibody-based protein measurements in yeast, which quantified a common tag engineered into each ORF (Supplementary Fig. 1a)[2]. The coverage of yeast ORFs was below two-thirds of the ORFeome for each of the two orthogonal data sets, suggesting that the proteome of yeast grown under standard laboratory conditions has been exhaustively mapped out by automated peptide sequencing or by antibody-based detection, and the two orthogonal data sets showed a high degree of overlap. Next, for each protein from PeptideAtlas we selected an optimal set of up to eight peptides with favourable mass-spectrometry properties and unique occurrence within the compiled protein sequence database (proteotypic peptides, PTPs)[16]. We predicted PTPs for proteins for which no empirical data were available (Supplementary Fig. 1b). For proteins that were not included in PeptideAtlas and that were not detected using the antibody-based methods, we selected at least two peptides with an isoelectric point below 4.5, if available, to maximize the probability of detecting the corresponding proteins if the peptide samples were first fractionated by off-gel electrophoresis[5]. Approximately 200 proteins remained refractory to these selection criteria, including proteins that do not generate any suitable tryptic peptide for mass-spectrometry analysis (Supplementary Fig. 1c). The final peptide set, comprising approximately 28,000 peptides, was synthesized on a small scale to assemble a peptide library representing 97% of the predicted yeast proteome (Fig. 1).

We next used these peptides to generate two reference spectral libraries, each one supporting a commonly used proteomic method. We analysed the peptide set on a linear ion trap (LIT)-type instrument (LIT–Orbitrap hybrid) to generate reference fragment ion spectra for spectral matching of data acquired in discovery experiments. We used a QQQ-type mass spectrometer (QTRAP hybrid) operated in the SRM-triggered tandem mass spectrometry (MS/MS) mode with fragmentation in the second quadrupole[17] to generate fragment ion spectra for the extraction of optimal SRM coordinates for targeted measurement of specific proteins (Fig. 1). The spectra acquired with each instrument were assigned to peptide sequences by sequence-database searching and the assignments were filtered to a false discovery rate (FDR) at a peptide-spectrum match level of less than 0.1% and at a peptide level of less than 0.5%, based on decoy counts.

To maximize proteome coverage, we combined our LIT data with quality-filtered LIT data from yeast extracts that had been submitted to PeptideAtlas[1] and consensus spectra from the National Institute of Standards and Technology (NIST) yeast ion-trap spectral library (http://peptide.nist.gov; build 19 October 2009). The QQQ data were

[1]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich CH-8093, Switzerland. [2]Department of Biology, Institute of Biochemistry, ETH Zurich, Zurich CH-8093, Switzerland. [3]Biotechnology Centre, Technische Universitaet Dresden, D-01069 Dresden, Germany. [4]Department of Chemical and Biomolecular Engineering and Division of Biomedical Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. [5]Institute for Systems Biology, Seattle, Washington 98103, USA. [6]Proteomics Core Facility Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland. [7]Biognosys AG, Zurich CH-8093, Switzerland. [8]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200031 Shanghai, China. [9]Max Planck Institute of Molecular Cell Biology and Genetics, D-01307 Dresden, Germany. [10]Competence Center for Systems Physiology and Metabolic Diseases, Zurich CH-8093, Switzerland. [11]Faculty of Science, University of Zurich, Zurich CH-8057, Switzerland.
*These authors contributed equally to this work.

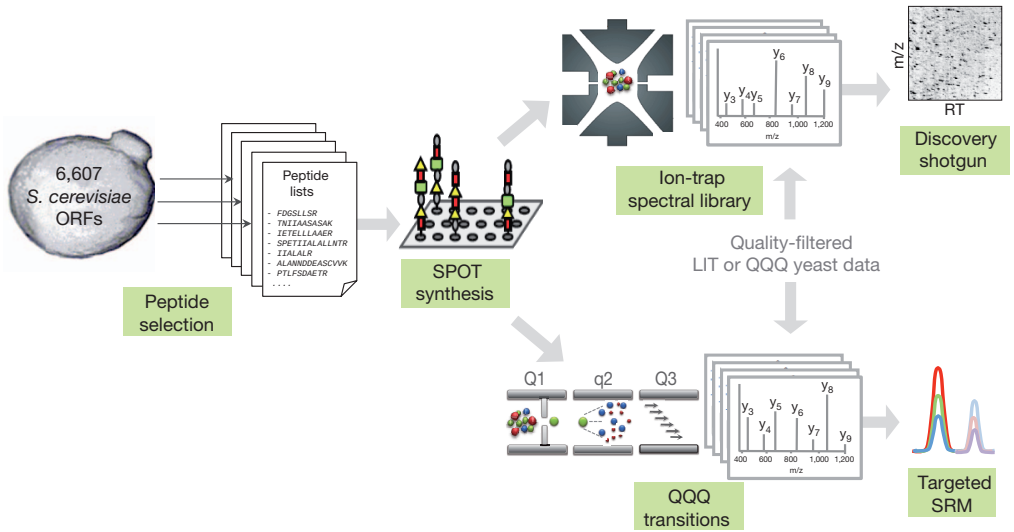

Figure 1 | Generation of a reference mass-spectrometric map for the yeast proteome. Schematic of the sequential steps of map generation: peptide selection based on the known ORFeome, small-scale peptide synthesis, QQQ- and LIT-type mass-spectrometry measurements, pooling of the generated spectra to existing data sets and application of the libraries to SRM measurements or spectral searches of yeast shotgun data sets. m/z, mass-to-charge ratio; Q1, first quadrupole of a triple-quadrupole mass spectrometer; q2, second quadrupole of a triple-quadrupole mass spectrometer (quadrupolar fragmentation cell); Q3, third quadrupole of a triple-quadrupole mass spectrometer; RT, retention time. On the SPOT synthesis membrane, different coloured shapes indicate different amino acids. Specific shapes and colours are not assigned to specific amino acids.

combined[11] with the data set acquired from yeast proteins from the original MRMAtlas[17]. For the assigned peptides, high-quality, de-noised consensus spectra were compiled[11]. The final LIT and QQQ spectral libraries contained consensus spectra for 100,815 and 28,216 peptide sequences, respectively, both covering 97% of the 6,607 sequences in the yeast proteome (Fig. 2a). For each peptide we then extracted from the QQQ spectral library the charge state (or states), information of the prominent fragment ions, including their masses, charges, relative intensities and chromatographic elution times, which collectively constitute a peptide SRM assay.

Our peptide selection criteria resulted in a peptide set that preferentially contained peptides of intermediate hydrophobicity (Fig. 2b). The approximately 1,630 synthetic peptides that could not be detected showed a bias towards extreme calculated[18] hydrophobicity values (Fig. 2b), indicating that very hydrophilic or very hydrophobic peptides are not well suited for chemical synthesis and/or liquid chromatography–mass spectrometry (LC–MS) analysis. Spectra acquired from synthetic peptides were indistinguishable from those acquired from the corresponding natural sources[17] (Supplementary Fig. 2).

The LIT and QQQ data sets were compiled into two 'builds' in PeptideAtlas[1], and can be downloaded or browsed interactively (Supplementary Discussion, Supplementary Figs 3–7 and http://www.srmatlas.org/yeast/).

We assessed the performance of the LIT library using a total yeast tryptic digest, analysed by data-dependent LC–MS/MS (data-dependent acquisition (DDA) mode). Spectral-library searching against the LIT library identified 1,617 unique proteins (Supplementary Fig. 8), compared to 1,529 proteins identified by conventional database searching[19], at the same protein FDR of 1%. Therefore, restricting the analysis to our chosen PTPs did not diminish the number of proteins identified. We next analysed the yeast sample multiple times, using distinct inclusion lists containing all precursor ions corresponding to the five top-ranking peptides per protein in the library. Spectral searching identified 2,509 unique proteins (Supplementary Figs 8 and 9). We then re-analysed the same sample using a single, reduced inclusion list containing only the peptide ions identified in the prior step (Supplementary Data set). Spectral matching against the LIT library identified 1,987 proteins (Supplementary Figs 8–10). This workflow resulted in approximately 30% more identifications than those achieved by conventional DDA followed by sequence-database searching. The spectral library search of this single file of about 34,000 MS/MS spectra took 5 min on one processor (approximately 0.01 s per spectrum), which implies that conducting on-the-fly data analysis is feasible in discovery-based proteomic experiments.

We next analysed the specificity of our SRM transitions in terms of uniqueness against two backgrounds of different complexity (Fig. 2c), using the SRMCollider tool[20] (Supplementary Discussion). Of the peptide precursors, 97.8% and 88.5% (Fig. 2c and http://www.srmatlas.org/yeast/) were predicted to be uniquely detected using the three highest transitions in our library, with and without time-scheduled acquisition, respectively, in the high complexity background. As simulations do not necessarily reflect the full complexity of a biological sample, it is advisable to measure four to five transitions per peptide and to use empirical relative intensities of fragment ions and peptide elution times as constraints to increase the assay specificity.

To demonstrate the utility of the two libraries in discovery and targeted proteomic experiments, we applied them to a protein-based quantitative trait locus (QTL) analysis in *S. cerevisiae*. Protein QTL studies aim to correlate protein abundance with genetic variation, and thus critically rely on the ability to measure protein concentrations precisely throughout large numbers of samples. Previous protein QTL studies[7] suffered from an inconsistent detection of peptides across samples (when a given peptide is present, it is not consistently detected) and a bias towards the detection of abundant proteins (Supplementary Discussion). To overcome these limitations we applied a two-step workflow based on our spectral libraries to a genetically diverse population of 78 yeast strains obtained by crossing a wild isolate (RM11-1a) and a strain isogenic to the standard S288c laboratory strain (BY4716)[21]. To identify proteins whose cellular concentrations are probably affected by QTLs in this cross, we carried out a discovery proteomic experiment on the two parental yeast strains and a subset of 16 segregants out of the total 78 strains that we selected, with the aim to maximize the genetic diversity between them (Supplementary Discussion and Supplementary Fig. 11). Using the single inclusion list described above and spectral matching against the LIT library, we identified approximately 2,500 uniquely mapped proteins at 1% FDR and quantified the abundance of the corresponding peptides throughout the different samples using a label-free approach. We excluded peptides with sequence differences between the RM11-1a and BY4716 backgrounds and proteins for which no high-quality

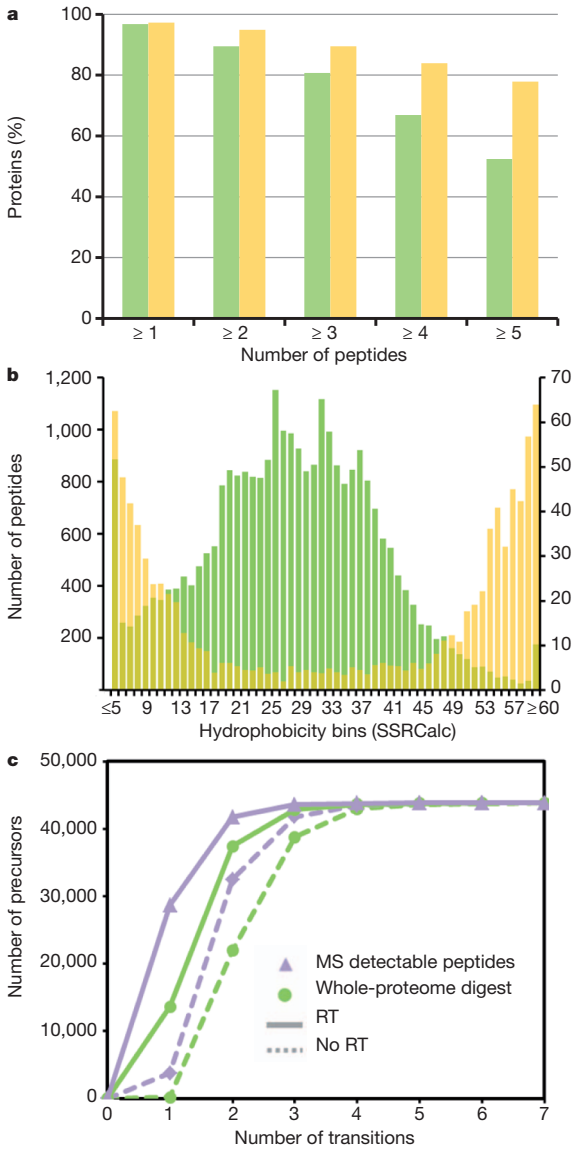

Figure 2 | Composition and use of the spectral libraries. **a**, Number of peptides per protein in the SRM assay (green) or ion-trap (yellow) library. The percentage is relative to the total of 6,607 theoretical *S. cerevisiae* proteins. **b**, Hydrophobicity distribution for all selected peptides (green, left axis) and for those that could not be detected with our QQQ workflow (yellow, right axis). Hydrophobicity is expressed as sequence-specific retention calculator (SSRCalc)[18] score bins. A low SSRCalc score indicates low hydrophobicity. **c**, Minimal number of transitions predicted to uniquely identify precursor ions in the QQQ library. A cumulative plot of the minimal number of transitions forming an assay (unique ion signature[20]) is shown for all the precursors in the QQQ library. Transitions were selected with decreasing intensity. Uniqueness for a set of transitions was established if no other peptide precursor produced transitions that contained all of the query transitions, using a tolerance of ± 0.35 Thomson (Th) for both precursor and fragment ions. For time-scheduled SRM acquisition, only co-eluting peptides were considered, based on predicted retention times[18] and a tolerance of ± 2.5 min. The analysis is shown for two different backgrounds: all yeast peptides potentially detectable by mass spectrometry, as derived from the yeast PeptideAtlas (purple), and all theoretical tryptic peptides in the yeast proteome (green). RT and no RT indicate whether time-scheduled SRM acquisition is considered.

ion features could be detected. We ranked the remaining approximately 2,100 proteins based on the variability of their concentrations across the sample strains, and corrected for the number of peptides that were measured per protein (Supplementary Fig. 12). To identify cellular processes and pathways that were particularly affected by genetic variation, we analysed the 150 proteins showing the highest variability for gene-ontology-based functional enrichment, metabolic-pathway enrichment, and protein module over-representation tests using protein-interaction data (Supplementary Tables 1–3). Proteins involved in NADH oxidation, arginine or ornithine biosynthesis, and amino acid metabolism were significantly ($P < 0.01$) enriched among the most variable proteins. We selected a set of 48 proteins that are members of the most highly enriched pathways and subnetworks. These proteins covered most levels of cellular abundances (Supplementary Fig. 13) and included a protein from one of the targeted pathways (Arg3) that was not detected in the discovery phase (Supplementary Table 4). We next used SRM assays (Supplementary Information) from the QQQ map for the targeted quantification of the 48 proteins throughout 76 segregants and both parental strains. The SRM-based quantification resulted in a highly consistent and comprehensive data set (Fig. 3a) and enabled the precise determination of inter-sample variations of protein abundances.

Epistatic interactions between genes are important factors contributing to the variation of complex traits and are thought to partly explain 'missing heritability' observed in traditional association studies[22]. However, detecting epistasis in QTL studies is notoriously difficult, mostly owing to lack of sufficient statistical power. We reasoned that because of the precision of SRM data and the relatively large number of samples it should be possible to detect epistasis. We therefore extended a machine-learning-based QTL mapping method that we developed previously[23,24] to also report epistatic interactions between pairs of loci affecting a common protein (Supplementary Figs 14–17). Application of this framework to our protein QTL data identified 32 protein QTLs involving single loci (FDR < 0.15) and 10 pairs of epistatic protein QTLs (FDR < 0.2) (Fig. 3b). In total, 28 of the 48 proteins were under the control of at least one protein QTL and 23 distinct genomic regions were involved. Thus, protein concentrations are strongly affected by natural genetic variation and epistatic interactions between loci affecting protein levels are a common phenomenon. A protein module consisting of Bat1, Bat2, Rpn11, Hsp60 and Ilv2 (which we termed the 'B1B2 module') caught our attention, because all of its components were involved in at least one significant protein QTL (Fig. 3b). The module functionally relates to protein turnover and amino acid metabolism, and is physically connected to mitochondria (Supplementary Discussion). Interestingly, six different genomic regions contained polymorphisms independently affecting the levels of the different proteins in the module. Detailed analysis revealed that segregant strains carrying the BY4716 alleles at the respective loci expressed consistently lower levels of all proteins that are part of this module, with Bat2 being the only exception (Fig. 3c). Importantly, Bat2 favours the reverse metabolic reaction catalysed by Bat1 (catabolism and anabolism of branched-chain amino acids, respectively, Supplementary Discussion)[25]. Thus, the two parental strains have acquired a set of independent genetic variations altering the abundance of B1B2 proteins such that the pathway activity consistently changes in one direction. Our data contained a second example of such coordinated acquisition of independent polymorphisms affecting the regulation of alcohol dehydrogenases (NADH module, Fig. 3b): Adh1, Adh3 and Adh5, which are upregulated in in the RM11-1a background, can catalyse the last reaction of the ethanol production pathway, whereas Adh2, which is linked to another locus and down-regulated in the RM11-1a background, catalyses the reverse reaction preferentially (Supplementary Fig. 18)[26,27].

We have described here a mass-spectrometric map for the near-complete proteome of yeast, to support both discovery-driven, and targeted proteomics experiments. The LIT library can be used for spectral matching of shotgun data, thus exploiting the previously acknowledged benefits of spectral matching, such as speed, confidence or an increase in the number of identifications. A limitation of the library is that it does not include protein post-translational modifications. However, in principle, the same approach can also be applied to generate spectral libraries for modified peptides and our libraries can

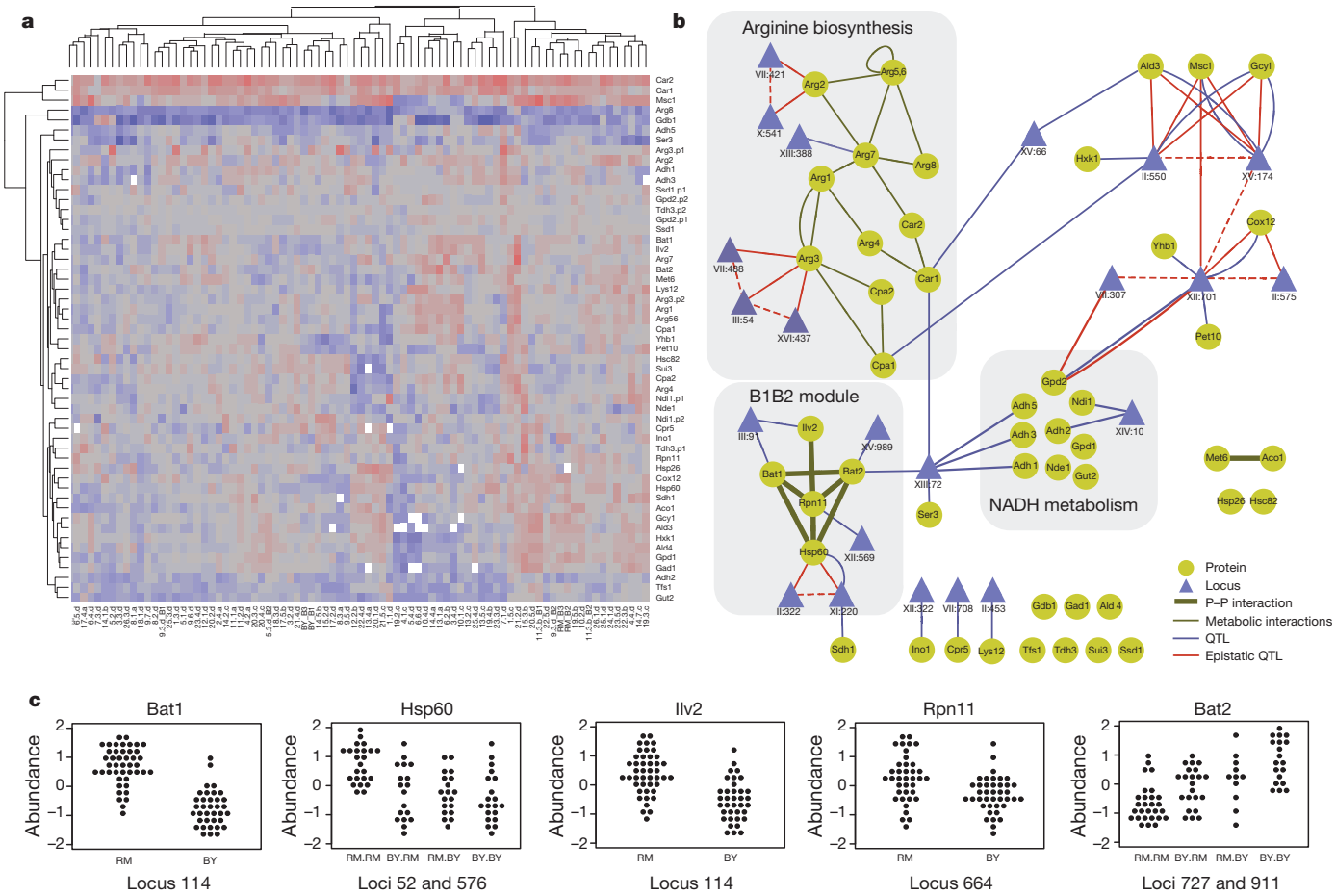

Figure 3 | Quantitative trait analysis from the targeted proteomic data set. **a**, Two-way cluster analysis of summarized protein abundances measured by SRM in 82 samples (78 strains) of the cross between RM11-1a (RM) and BY4716 (BY). Columns are clustered according to the samples, and rows are clustered according to the proteomic traits. Abundance levels are colour coded in a blue–red gradient (blue, low abundance; red, high abundance; white, missing data). The completeness of the data set reaches 99.5% even though the summarization procedure was conservative and generated additional missing values. **b**, Network representation of protein abundance QTL for the 48 targeted proteins. Locus positions are indicated by the chromosome name followed by the genomic position of the centre of the locus in kilobases. Physical protein–protein interactions (P–P interaction) were obtained from BioGRID (http://thebiogrid.org). Metabolic interactions were manually reconstructed from BioCyc (http://biocyc.org/). Epistatic protein QTL always link two loci (connected by a dashed line) that are in epistasis with respect to a protein abundance trait (connected by a solid line). **c**, Protein abundances for the B1B2 module. For each protein, abundances are shown for groups of strains separated according to their genotype at the respective protein QTL. A pair of interacting loci was linked to the variations of Hsp60. The epistatic interaction is clearly visible when the RM11-1a allele is inherited at both loci. In contrast, the effects of the two loci linked to Bat2 are additive. Notably, the directionality of regulation is shared by all components of the module except Bat2 (overexpression when the RM11-1a alleles are inherited). Bat1 and Bat2 are paralogues catalysing the same metabolic reaction in opposite directions. RM.RM, BY.RM, RM.BY and BY.BY indicate allele combinations at two interacting loci.

be iteratively expanded and enriched to improve their coverage and quality. The library of SRM coordinates substantially expands the capabilities of SRM-based targeted proteomic experiments and accelerates their implementation. The availability of QQQ fragment-ion relative intensities and peptide elution times can be used for the automated scoring and statistical evaluation of large-scale SRM data sets with respect to their FDR[28]. A limitation of our SRM assay library is that we do not experimentally determine the sensitivity and specificity of each assay. These properties are sample- and platform-dependent (for example, chromatography, resolution and tuning) and should therefore be determined locally, for any particular sample. In addition, as our reference spectra support the selection of multiple intense transitions for each peptide, one or more can be discarded if found to be locally unspecific. For example, although the peptides that differed between the parental backgrounds were discarded in our protein QTL study, 90% of the SRM library remained usable despite 1% divergence between the reference and the RM11-1a genomes.

The libraries presented here are also a useful blueprint for studying peptide-fragmentation properties in quadrupoles and ion-trap mass spectrometry and will support the development of new acquisition methods relying on the knowledge of peptide fragmentation patterns. A first example of such a data-independent acquisition method, SWATH (sequential windowed acquisition of all theoretical fragment-ion spectra) mass spectrometry, has been described already[29].

The application of our libraries to yeast protein QTL analysis resulted in the precise measurement of proteins spanning a broad range of abundances (Supplementary Figs 13 and 19) throughout a large number of samples. This enabled the detection of novel protein QTLs and epistatic interactions, and led to the identification of two cases of co-inheritances of several independent genetic variations that influence the abundance of related proteins in a biologically coherent manner. A deeper analysis of these cases together with previous findings[30] suggests that they are examples for the adaptive evolution of protein-level regulation (Supplementary Discussion).

The proteome map generated here, the publicly accessible tools to navigate it, and the data acquisition and processing strategies that are enabled by such resources expand the capabilities of current proteomics experiments and substantially improve their performance features.

METHODS SUMMARY

We selected a set of peptides for each of the 6,607 proteins in the *S. cerevisiae* proteome based on existing proteomic data sets or computational prediction. We synthesized the peptide set using SPOT synthesis and analysed the crude peptides in batches on both a QQQ and a LIT mass spectrometer. We pooled the generated spectra with spectra extracted from public repositories. The resulting QQQ and LIT spectral libraries were made publicly available through the PeptideAtlas interface (http://www.srmatlas.org/yeast/; Supplementary Methods). We used the libraries to analyse a collection of yeast strains grown in glucose-based medium. Proteins were identified and quantified using a shotgun and a targeted proteomic approach, both guided by the mass spectrometric coordinates contained in the library. Protein QTL mapping was carried out using a modified Random Forest machine learning method that also reports pairs of epistatically interacting loci. FDRs of protein QTLs were obtained by permuting the phenotype vectors 25,000 times and thereby retrieving an individual background score distribution for each genetic marker.

Full methods and supplementary material accompany this paper.

Received 3 February 2011; accepted 30 November 2012.
Published online 20 January 2013.

1. Deutsch, E. W. The PeptideAtlas Project. *Methods Mol. Biol.* **604,** 285–296 (2010).
2. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).
3. de Godoy, L. M. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455,** 1251–1254 (2008).
4. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
5. Picotti, P. *et al.* Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138,** 795–806 (2009).
6. Anderson, L. & Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5,** 573–588 (2006).
7. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39,** 1369–1375 (2007).
8. Taussig, M. J. *et al.* ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature Methods* **4,** 13–17 (2007); erratum **4,** 187 (2007).
9. Deutsch, E. W., Lam, H. & Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9,** 429–434 (2008).
10. Lam, H. & Aebersold, R. Spectral library searching for peptide identification via tandem MS. *Methods Mol. Biol.* **604,** 95–103 (2010).
11. Lam, H. *et al.* Building consensus spectral libraries for peptide identification in proteomics. *Nature Methods* **5,** 873–875 (2008).
12. Craig, R. *et al.* Using annotated peptide mass spectrum libraries for protein identification. *J. Proteome Res.* **5,** 1843–1849 (2006).
13. Lange, V. *et al.* Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4,** 222 (2008).
14. Lange, V. *et al.* Targeted quantitative analysis of *Streptococcus* pyogenes virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* **7,** 1489–1500 (2008).
15. Ahrens, C. H. *et al.* Generating and navigating proteome maps using mass spectrometry. *Nature Rev. Mol. Cell Biol.* **11,** 789–801 (2010).
16. Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nature Biotechnol.* **25,** 125–131 (2007).
17. Picotti, P. *et al.* A database of mass spectrometric assays for the yeast proteome. *Nature Methods* **5,** 913–914 (2008).
18. Krokhin, O. V. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. *Anal. Chem.* **78,** 7785–7795 (2006).
19. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467 (2004).
20. Röst, H., Malmstrom, L. & Aebersold, R. A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol. Cell. Proteomics* **11,** 540–549 (2012).
21. Brem, R. B. *et al.* Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755 (2002).
22. Zuk, O. *et al.* The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109,** 1193–1198 (2012).
23. Michaelson, J. J. *et al.* Data-driven assessment of eQTL mapping methods. *BMC Genomics* **11,** 502 (2010).
24. Ackermann, M., Clément-Ziza, M., Michaelson, J. J. & Beyer, A. Teamwork: improved eQTL mapping using combinations of machine learning methods. *PLoS ONE* **7,** e40916 (2012).
25. Colón, M. *et al. Saccharomyces cerevisiae* Bat1 and Bat2 aminotransferases have functionally diverged from the ancestral-like *Kluyveromyces lactis* orthologous enzyme. *PLoS ONE* **6,** e16099 (2011).
26. Young, E. T. & Pilgrim, D. Isolation and DNA sequence of ADH3, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **5,** 3024–3034 (1985).
27. de Smidt, O., du Preez, J. C. & Albertyn, J. Molecular and physiological aspects of alcohol dehydrogenases in the ethanol metabolism of *Saccharomyces cerevisiae. FEMS Yeast Res.* **12,** 33–47 (2012).
28. Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nature Methods* **8,** 430–435 (2011).
29. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Proteomics* **11,** O111–016717 (2012).
30. Fraser, H. B., Moses, A. M. & Schadt, E. E. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl Acad. Sci. USA* **107,** 2977–2982 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project has been funded in part by ETH Zurich, the Swiss National Science Foundation (3100A0-107679), the National Heart, Lung and Blood Institute, National Institutes of Health (N01-HV-28179), the National Science Foundation MRI (grant 0923536), the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg, and by SystemsX.ch, the Swiss initiative for systems biology. P.P. is supported by a Foerderungsprofessur grant from the Swiss National Science Foundation (PP00P3_133670), by a European Union Seventh Framework Program Reintegration grant (FP7-PEOPLE-2010-RG-277147) and by a Promedica Stiftung (2-70669-11), H. L. is supported by the University Grant Council of the Hong Kong Special Administrative Region Government, China (HKUST DAG08/09.EG02). A.B. is supported by the Klaus Tschira Foundation and by a European Union FP7 HEALTH grant (HEALTH-F4-2008-223539). R.A. is supported by the European Research Council (ERC-2008-AdG 233226) and by SystemsX.ch, the Swiss Initiative for Systems Biology.

Author Contributions P.P. and M.C.-Z. carried out the experiments; M.C.-Z., H.L., E.W.D., O.R., L.R., P.P., J.J.M., A.B. and R.A. conceived the data analysis pipeline; P.P., M.C.-Z., H.L., D.S.C., A.S., E.W.D., H.R., Z.S., O.R., L.R., J.J.M. and Q.S. analysed the data; A.S., A.F., Q.S. and U.K. performed mass-spectrometry measurements; P.P., A.B., M.C.-Z., S.A. and R.A. designed the experiments; P.P., M.C.-Z., A.B., H.L., H.R., S.A. and R.A. wrote the manuscript; B.W. and R.L.M. supervised part of the project; and R.A., A.B. and P.P. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A. (aebersold@imsb.biol.ethz.ch), A.B. (andreas.beyer@biotec.tu-dresden.de) or P.P. (paola.picotti@bc.biol.ethz.ch).