

An integrated workflow for charting the human interaction proteome: insights into the PP2A system

Timo Glatter^{1,2,5}, Alexander Wepf^{1,2,5}, Ruedi Aebersold^{1,2,3,4} and Matthias Gstaiger^{1,2,*}

¹ Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland, ² Competence Center for Systems Physiology and Metabolic Diseases, ETH Zurich, Zurich, Switzerland, ³ Faculty of Science, University of Zurich, Zurich, Switzerland and ⁴ Institute for Systems Biology, Seattle, WA, USA

⁵ These authors contributed equally to this work

* Corresponding author. Institute of Molecular Systems Biology, ETH, Wolfgang Pauli Strasse 16, Zürich 8093, Switzerland.

Tel.: +41 44 633 71 49; Fax: +41 44 633 10 51; E-mail: gstaiger@imsb.biol.ethz.ch

Received 26.6.08; accepted 4.12.08

Protein complexes represent major functional units for the execution of biological processes. Systematic affinity purification coupled with mass spectrometry (AP-MS) yielded a wealth of information on the compendium of protein complexes expressed in *Saccharomyces cerevisiae*. However, global AP-MS analysis of human protein complexes is hampered by the low throughput, sensitivity and data robustness of existing procedures, which limit its application for systems biology research. Here, we address these limitations by a novel integrated method, which we applied and benchmarked for the human protein phosphatase 2A system. We identified a total of 197 protein interactions with high reproducibility, showing the coexistence of distinct classes of phosphatase complexes that are linked to proteins implicated in mitosis, cell signalling, DNA damage control and more. These results show that the presented analytical process will substantially advance throughput and reproducibility in future systematic AP-MS studies on human protein complexes.

Molecular Systems Biology 20 January 2009; doi:10.1038/msb.2008.75

Subject Categories: proteomics; signal transduction

Keywords: human interactome; mass spectrometry; PP2A; protein complexes

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. Creation of derivative works is permitted but the resulting work may be distributed only under the same or similar licence to this one. This licence does not permit commercial exploitation without specific permission.

Introduction

The majority of proteins exert an effect in the context of macromolecular assemblies that are part of dynamic networks of enormous complexity. Cellular processes, such as cell signalling, proliferation, apoptosis and cell growth, emerge to a large extent from the properties of such networks. Hence, understanding and modelling of cellular processes in healthy and pathological conditions depend on comprehensive and robust information on the topology and the dynamic properties of the engaged protein networks. Initially, large-scale protein interaction studies were performed with the yeast two-hybrid technology, which provided insights into global patterns of binary protein interactions of model organism proteomes (Uetz *et al.*, 2000; Walhout *et al.*, 2000; Ito *et al.*, 2001). More recently, affinity purification coupled with mass spectrometry (AP-MS) has become the method of choice for the analysis of protein complexes under near-physiological conditions (Gingras *et al.*, 2007; Kocher and Superti-Furga, 2007). Large-scale AP-MS studies performed in yeast provided the first comprehensive set of high-density interaction data,

which became an invaluable source of information for yeast systems biology (Gavin *et al.*, 2002, 2006; Ho *et al.*, 2002; Krogan *et al.*, 2006). The success in yeast can be mainly attributed to the high efficiency of homologous recombination that allowed genome-wide tagging of yeast ORFs as a valuable resource for large-scale AP-MS studies. However, no such genetic system exists for multicellular eukaryotes.

Given the various cell types and cellular states, each characterized by specific protein-protein interaction networks, the complexity of the human proteome and the limited genetic methods available to generate cell lines expressing affinity-tagged proteins, global analysis of protein complexes and protein interaction networks in human cells is a daunting task. Progress towards this goal will strongly depend on efficient and robust AP-MS workflows for human cells that provide comprehensive as well as high-confidence protein complex information to populate public databases. The robustness and reproducibility of such methods are key because it can be anticipated that data from different studies and research groups must be combined to achieve saturation coverage of the human interaction proteome. However, false

discovery and reproducibility rates are not known for existing methods, which make the combination of AP-MS data from different studies difficult. In addition, present AP-MS strategies are limited by the labour-intensive generation of large collections of human cell lines for expression of epitope-tagged bait proteins, the low yield of protein complex isolation from such cell lines and the limited sensitivity of MS-based protein identification.

To overcome some of these major limitations, we have developed an integrated experimental workflow. Besides optimizing each experimental step, we focused on the compatibility of the steps with each other to generate a process with improved performance. As a result, the proposed procedure significantly enhanced the throughput of generating bait-expressing cell lines, increased the protein complex purification yields by a novel double-affinity strategy and allowed analysis of protein complexes and interaction networks with high sensitivity and reproducibility. We applied this procedure to study a network of human protein phosphatase 2A (PP2A) complexes. PP2A is a heterotrimeric, evolutionary conserved serine/threonine phosphatase with regulatory functions in a wide range of cellular processes, including transcription, apoptosis, cell growth and cellular transformation (Virshup, 2000; Lechward *et al*, 2001). The human genome encodes two catalytic subunits (PPP2CA, PPP2CB), two scaffolding subunits (PPP2R1A, PPP2R1B) and at least 15 known regulatory B subunits that, by combinatorial assembly, can potentially form a multitude of different trimeric PP2A complexes (Janssens and Goris, 2001; Lechward *et al*, 2001). It is believed that the versatile nature of this combinatorial subunit arrangement provides substrate specificity as well as temporal and spatial control of phosphatase activity. However, no systematic study has yet been performed to address the question, which PP2A complex forms indeed coexist in human cells and how these complexes are connected to specific cellular processes through protein-protein interactions. Using the method described in this work, we identified 197 specific protein-protein interactions at a reproducibility rate of at least 85%. The discovered interactions constitute a network of different classes of concurrently present phosphatase complexes that in turn are linked to proteins with specific functions in cell signalling, mitosis, DNA repair and more.

On the basis of these results, we believe that the proposed analytical procedure will significantly improve the scope and reproducibility of future AP-MS studies on the human interaction proteome.

Results

An integrated workflow for systematic AP-MS studies on human protein complexes

Affinity purification coupled with mass spectrometry analysis of protein complexes can be grouped into three sequential steps: production of cell lines expressing epitope-tagged bait proteins, protein complex purification and MS-based analysis of the isolated samples. Each step contributes to the overall performance of the process. To generate a robust and reproducible workflow for the characterization of human

protein complexes, we have optimized each step and integrated them into an efficient process. Thereby, we paid attention to a good compatibility of the steps between each other. The system builds on (i) gateway-compatible orfeome collections and the Flippase (Flp) recombination system to rapidly generate large collections of human cell lines by homologous recombination for isogenic and tetracycline (tet)-controlled expression of tagged bait proteins, (ii) the development of a novel double-affinity purification strategy to significantly increase sample recovery and reproducibility and (iii) direct liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis of purified complexes to improve the sensitivity of protein identification. In what follows we describe the individual steps of the workflow and document its performance for systematic protein complex analysis as shown for the human PP2A protein interaction network.

Rapid generation of cell lines for inducible expression of affinity-tagged bait proteins

A major bottleneck in large-scale AP-MS analyses in species other than *Saccharomyces cerevisiae* has been the resource-intensive generation of cell lines expressing epitope-tagged bait proteins, preferably at controlled levels, required for protein complex purification. Here, we combined recombinational cloning of expression constructs from human orfeome libraries with homologous recombination using Flp recombinase in human cells to significantly increase the production rate for such human cell lines (Figure 1A). We used a gateway-compatible orfeome collection containing 12 212 ORFs, representing 10 214 non-redundant protein-coding genes (Lamesch *et al*, 2007) as a resource to generate expression constructs by LR recombination with a destination vector suitable for tetracycline-controlled expression of affinity-tagged bait proteins. The presence of a Flp recombination target site (FRT) in the resulting expression constructs supported the rapid generation of bait-expressing cell lines by Flp-mediated recombination with a single FRT site present in the HEK293 host cell line (O’Gorman *et al*, 1991). The system was evaluated with respect to the following properties. (i) Efficiency and reliability. We routinely obtained isogenic human HEK293 cell pools within 2 weeks after transfection with a success rate of about 85% ($n > 200$ transfected orf clones, data not shown) without further need for subcloning. (ii) Uniformity. The FRT recombination system ensures uniform expression of the transgene in the respective cell populations as demonstrated by indirect immunofluorescence microscopy of HEK293 cell lines expressing different epitope-tagged proteins from the human PP2A system (Figure 1B). (iii) Inducible bait expression. The ability to control bait protein expression levels is crucial in cases in which growth inhibitory or pro-apoptotic bait proteins are expressed. Expression levels of bait proteins could be induced and adjusted by tetracycline (Figure 1C) and were comparable with corresponding endogenous protein levels (Figure 1D). In conclusion, orfeome-based generation of human cell lines using the FRT system is an efficient and reliable method for the generation of large collections of isogenic cell pools for tetracycline-controlled expression of affinity-tagged bait proteins at close to physiological levels.

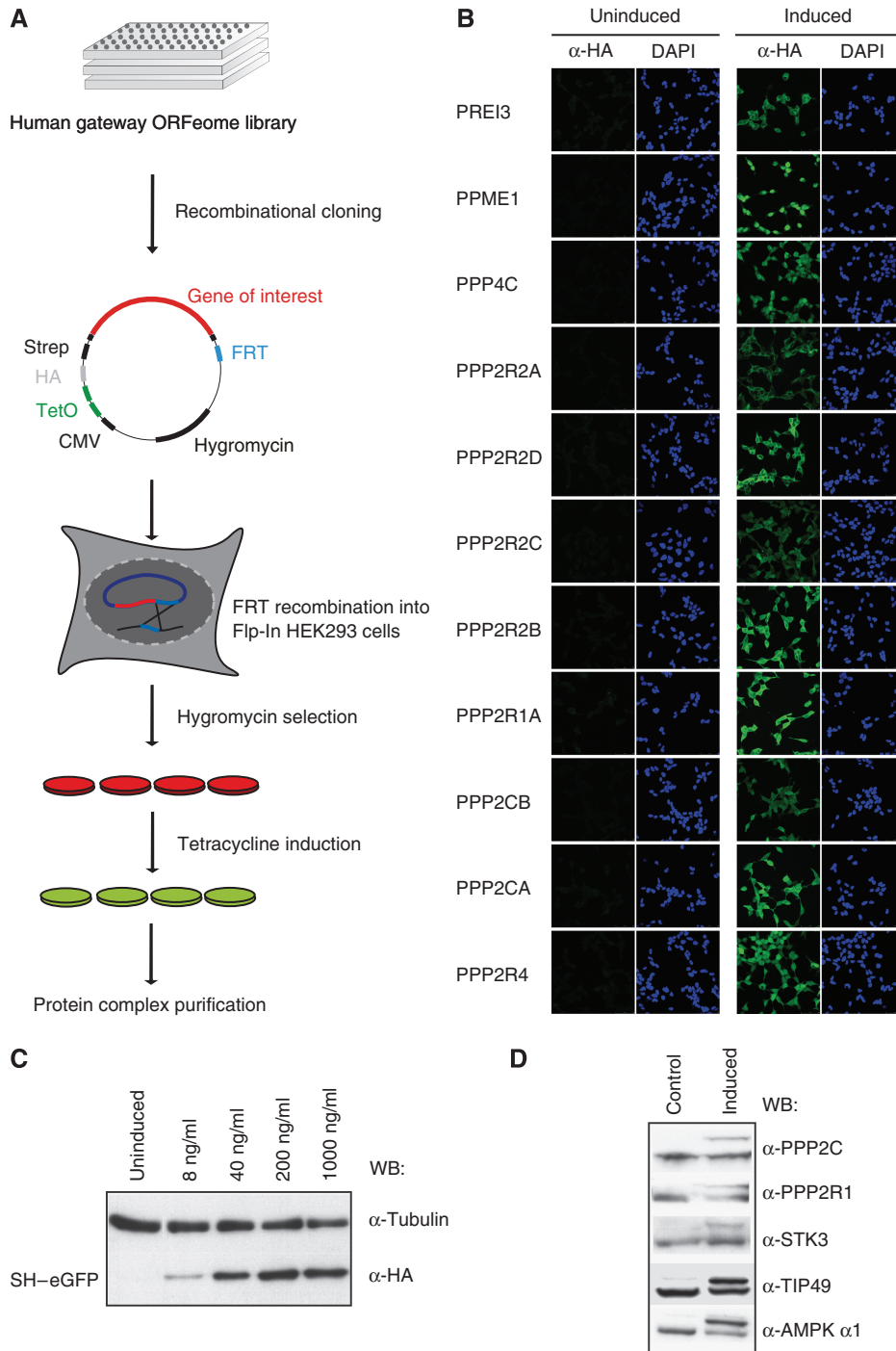


Figure 1 Rapid generation of cell line collections for isogenic and inducible bait expression using Flp-recombinase-mediated recombination. **(A)** Schematic overview on the generation of cell line collections. Starting from human Gateway orfeome collections, cDNAs of interest are recombined into an expression construct for tetracycline (tet)-inducible expression of strep-hemagglutinin double-tagged (SH) bait proteins. Isogenic cell lines are generated using Flp-recombinase-mediated recombination through single FRT sites present in the expression construct and the genome of Flp-In HEK293 cells stably expressing the tet repressor. After transfection, cell lines are selected on hygromycin for 2 weeks, tested for tetracycline-inducible expression and used for subsequent affinity purification. **(B)** Isogenic bait protein expression in HEK293 Flp-In cells. The expression of the indicated recombinant proteins in the absence or presence of 1 μ g/ml tetracycline for 24 h was visualized by indirect fluorescence microscopy with an anti-HA antibody. Nuclei were stained with DAPI. **(C)** Tet-inducible bait expression. Increasing amounts of tetracycline were added to HEK293 cells expressing SH-eGFP for 24 h. Bait expression was monitored by immunoblotting using anti-HA antibodies. **(D)** Comparison of protein expression levels of SH-tagged bait proteins with endogenous protein levels. HEK293 cell lines expressing SH-tagged bait proteins were analysed by immunoblotting using the indicated antibodies following induction with tetracycline (1 μ g/ml) for 24 h. HEK293 cells that do not express the corresponding bait proteins were used as controls. Note that anti-PPP2C and anti-PPP2R1 antibodies do not distinguish between the highly related endogenous proteins PPP2CA and PPP2CB or PPP2R1A and PPP2R1B, respectively.

Increased yields in protein complex preparations by SH-double affinity purification

So far, tandem affinity purification (TAP) has been the most widely used procedure in systematic protein complex purification (Rigaut *et al*, 1999; Bouwmeester *et al*, 2004; Gingras *et al*, 2005; Gavin *et al*, 2006). However, these studies required large amounts of cellular starting material due to the low purification yields obtained by the TAP procedure (Al-Hakim *et al*, 2005; Gregan *et al*, 2007). This imposes significant economical and logistic challenges, especially for large-scale studies. To improve the yield, we integrated a novel double-affinity purification protocol into our analytical workflow. Protein complexes are isolated through a small double-affinity tag (SH-tag) consisting of a streptavidin-binding peptide and a hemagglutinin (HA) epitope tag. In addition, we optimized the

purification protocols for efficient double-affinity purification from low amounts of starting material. Following induction of isogenic bait expression using tetracycline, SH-tagged and associated proteins are first bound to an affinity column containing a modified version of streptavidin (Junttila *et al*, 2005) and specifically eluted with biotin onto an anti-HA antibody column. The protein complexes are eluted from this column at low pH (Figure 2A). Western blotting showed that SH-PPP2R2B from HEK293 cell extracts could be bound near quantitatively to the streptavidin column (Figure 2B, SNS). Elution from the streptavidin column with biotin was highly efficient with almost no detectable bait protein left on the streptavidin beads following elution with SDS Laemmli buffer (not shown). Overall, the first purification step recovered more than 90% of bait protein (Figure 2B, ES). From the western blot signal intensity of the final eluate (Figure 2B, EH), we estimate

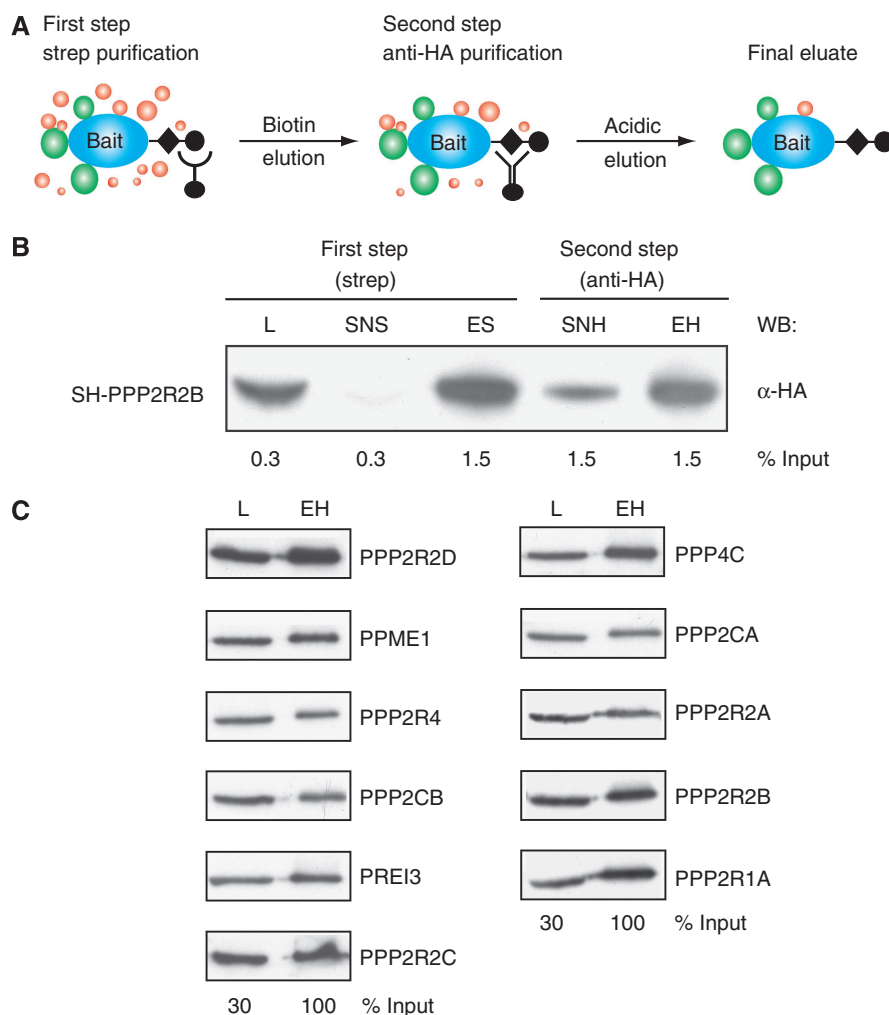


Figure 2 Monitoring SH-double-affinity purification efficiency. **(A)** Schematic overview of the purification procedure. HEK293 cells expressing SH-tagged proteins are lysed and first purified from total protein extracts using streptavidin sepharose (Strep-Tactin beads). After several wash steps, purified proteins are released in the presence of 2 mM biotin for subsequent immunoaffinity purification using anti-HA agarose. Finally, protein complexes are eluted with 0.2 M glycine, pH 2.5, and processed for mass spectrometric analysis. **(B)** Western blot analysis of SH-purification yields. SH-PPP2R2B expressing HEK293 cell line was generated as described in Figure 1 and the purification was performed on 3×10^7 cells. The purification procedure was monitored by immunoblotting using anti-HA antibodies. L: lysate; SNS: supernatant after streptavidin purification; ES: elution from the streptavidin sepharose; SNH: supernatant after anti-HA purification; EH: elution from the anti-HA agarose (final eluate). Information on the percentage of input is given to compare the relative amount of sample loaded on each gel lane. **(C)** Reproducibility of SH-purification yields. 11 cell lines inducibly expressing SH-tagged bait proteins related to the PP2A phosphatase system were generated as described. Lysate (L) and final eluate (EH) from all 11 bait-specific SH-purifications were immunoblotted using anti-HA antibodies and percentage of loaded sample amount is indicated.

the overall yield of the double purification at about 30–40% of bait protein present in the cell lysate, which is among the highest reported for double-affinity purification protocols. Importantly, the high yields were achieved independent of the bait protein, as SH-purifications of 11 different bait proteins showed comparable yields (Figure 2C).

Monitoring specificity and sensitivity of the SH-purification

Protein complex preparations typically contain significant amounts of co-purifying contaminant proteins that increase sample complexity, reduce the sensitivity of detection of true interactors and complicate the interpretation of the results. We first monitored the sample complexity during the SH-purification step of our workflow by silver staining after biotin (ES) and final elution (EH) using the PP2A regulatory B subunit PPP2R2B as a bait (Figure 3A). Biotin eluates contained high molecular weight contaminants, which may interfere with the identification of less-abundant interaction partners by direct LC-MS/MS. These contaminants were efficiently removed by applying the second purification step as shown by silver staining and direct LC-MS/MS (Figure 3A, Supplementary Table I).

In addition, we used the recently developed MasterMap concept to illustrate the specificity increase achieved by the

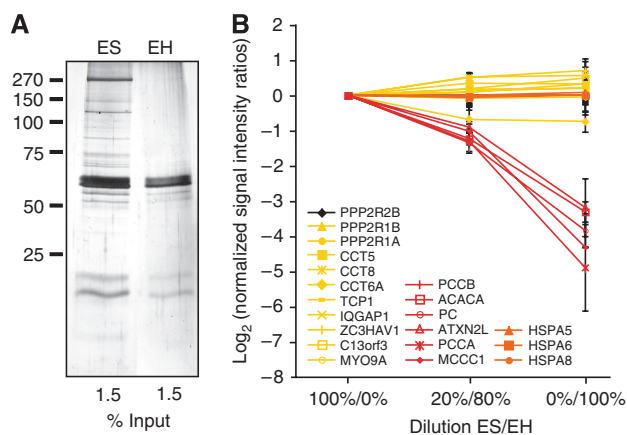


Figure 3 Analysis of SH-double-affinity purification specificity. **(A)** Silver stain monitoring of sample complexity. SH-double-affinity purification was performed on 3×10^7 HEK293 cells expressing SH-PPP2R2B, and aliquots of the first (ES) and second (EH) purification step were separated by SDS-PAGE. **(B)** Quantitative MS analysis of the specificity increase by the second purification step. HEK293 SH-PPP2R2B lysates were split in half. Single and double-affinity purifications were performed as described before. Peptides derived from single (ES) and double-affinity purification (EH) were mixed in a three-step dilution. Following MS analysis, MS spectra were aligned and MS1 signal intensities were used for relative quantification to generate protein abundance profiles across the three samples. All presented protein profiles were generated from aligned MS1 features that correspond to at least five unmodified, fully tryptic peptides. Note that not all observed protein profiles were included here, as they did not pass the indicated filtering criteria (e.g. PPP2CA). Protein abundance profiles were normalized to the bait profile. Yellow lines correspond to specific binding partners that match the bait profile. Orange lines refer to unspecifically co-purifying proteins identified also in SH-eGFP control samples. Profiles in red represent proteins unspecifically co-purified with the streptavidin sepharose beads that are successfully removed after the second purification step. Error bars indicate s.e.m. of MS1 feature ratios of the indicated proteins.

second purification step by monitoring the relative abundance profiles of co-purifying proteins (Rinner *et al*, 2007). This approach allows label-free protein quantification from aligned MS1 spectra obtained on a high mass accuracy instrument. MS1 quantification of the proteins identified with at least five fully tryptic unmodified peptides revealed three major groups of co-purifying proteins. (i) One set of proteins (Figure 3B, red lines) was efficiently removed by the second purification. These proteins were also identified in SH-eGFP control samples (Supplementary Table II) and include a group of abundant carboxylases (e.g. MCCC1, PCCA, ACACA) that are known to be biotinylated *in vivo* (Gravel and Narang, 2005), and hence most likely interact with the streptavidin column independent of the bait protein. (ii) A group of proteins including HSP70 chaperones (HSPA5, HSPA6 and HSPA8) that remained in the sample even after the second purification step (Figure 3B, orange lines). These proteins most likely represent unspecific interactors that bind independently of the bait protein, as they were also identified in eGFP control experiments (Supplementary Table II). (iii) Finally, the group of specific interactors that followed the profile of the bait protein but were absent in SH-eGFP control purifications. This group contained well-established interactors of the bait protein PPP2R2B, including PPP2R1A and PPP2R1B (Figure 3B, yellow lines).

Previous systematic studies were hampered by the large amount of cellular starting material required for AP-MS analysis. We performed SH-purifications from as low as 4×10^6 HEK293 cells. Direct LC-MS/MS analysis of 25% of the tryptic digest was still sufficient to identify PPP2R2B-interacting proteins PPP2CA, PPP2R1A and most of the associated subunits of the CCT complex (Supplementary Table III). As PP2A is regarded as an abundant phosphatase, we recommend to use 3×10^7 cells for standard SH-purification of protein complexes.

Conclusively, the SH-double-affinity purification step is a central part of our workflow and results in protein complex preparations of high purity and, when combined with direct LC-MS/MS, it significantly reduces the amounts of cellular starting material required for large-scale analysis.

Direct LC-MS/MS analysis

After cell line generation and affinity purification, the final MS analysis step represents another experimental bottleneck in large-scale studies on protein complexes. Previous AP-MS workflows used SDS-PAGE fractionation before MS analysis (Gavin *et al*, 2006; Ewing *et al*, 2007). This, however, multiplies the workload and time required for MS analysis, as each band has to be analysed by an individual MS run. Direct LC-MS/MS has poorly been exploited in previous high-throughput AP-MS studies but can significantly enhance sensitivity and reduce instrument time of an AP-MS workflow, as the entire affinity-purified sample can be analysed by a reversed-phase liquid chromatography unit coupled to a mass spectrometer in one single step. For successful direct LC-MS/MS analysis, samples have to be of moderate complexity, meaning that the affinity-purified protein complexes should contain a minimum of contaminating proteins. This is necessary because direct LC-MS/MS is prone to an under-

sampling of MS1 features for fragmentation, an effect that is more pronounced in complex samples (Liu *et al*, 2004). Another prerequisite is the chemical compatibility of the sample with the subsequent LC-MS/MS step. We adapted the purification strategy specifically to address these two points. As described already above, SH-double-affinity purification results in pure samples preparations with very little contaminant protein present. In addition, the second purification step efficiently removes detergent, protease inhibitors and eluting reagents (e.g. biotin) present in the sample, which would interfere with subsequent LC-MS/MS. Following tryptic digest, the samples are desalted by a reversed-phase chromatography step and directly loaded on a reversed-phase HPLC column attached to an MS instrument for peptide separation and MS analysis. The entire LC-MS analysis is completed within 90 min and thus significantly increases sample throughput.

Reproducibility of AP-MS data sets

Reproducibility of the data obtained by AP-MS is key to assess the quality of protein interaction data sets but is poorly addressed for existing AP-MS workflows performed in human cells. We studied the overall reproducibility of protein interaction data obtained by the method described here at the network level. For this, we selected 11 ORF clones encoding proteins previously linked to the PP2A system and generated isogenic HEK293 cell lines for their expression as SH-tagged bait proteins (Figure 4A, upper panel). From each of these cell lines, we performed two biological replicate SH-purification experiments and analysed each sample once by LC-MS/MS. We identified a total of 185 proteins with at least one unique peptide and a ProteinProphet probability > 0.9 (Supplementary Table IV) (Nesvizhskii *et al*, 2003), which could be mapped to 165 unique human entrez gene IDs. To eliminate co-purifying contaminant proteins, we generated a database of

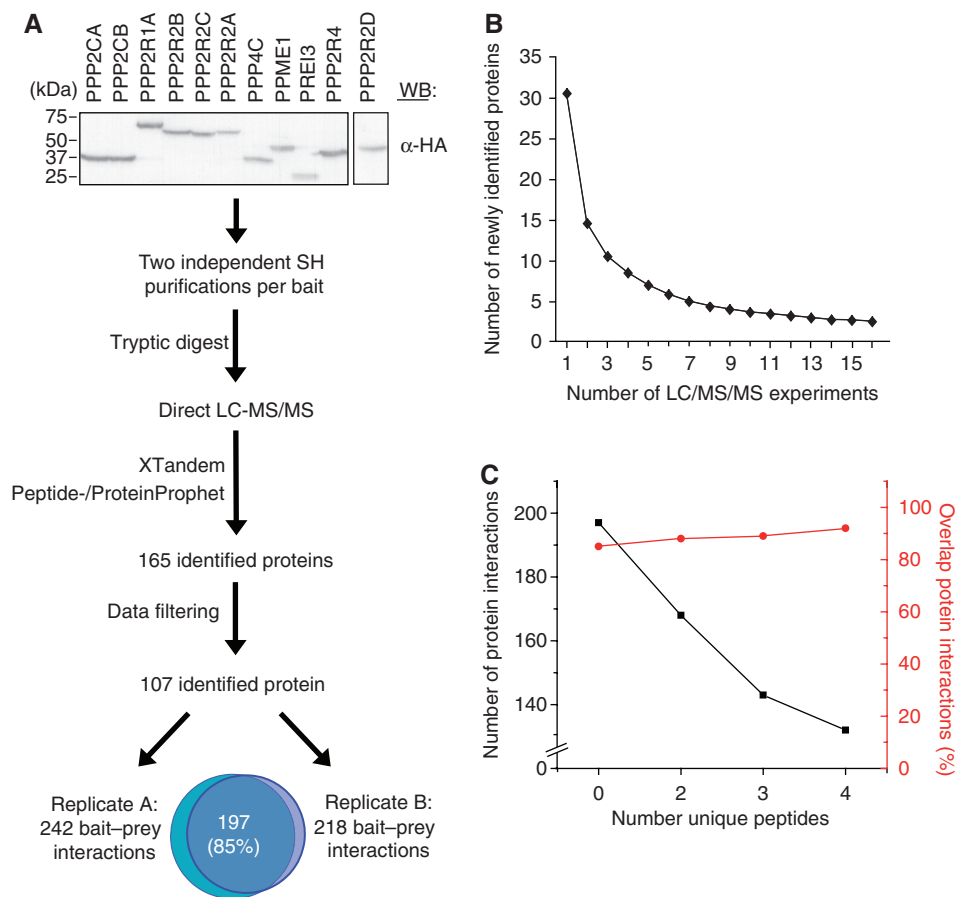


Figure 4 Data processing and reproducibility of the overall workflow for a human PP2A protein interaction network. **(A)** HEK293 cell lines expressing 11 different bait proteins linked previously to the human PP2A network were generated as described in Figure 1. Inducible expression was tested by western blotting (upper panel). Eleven cell lines were used for two independent SH-double affinity purification experiments directly followed by LC-MS/MS (SH-LC-MS/MS). Obtained mass spectra were analysed with XTandem followed by statistical validation of the search results using PeptideProphet and ProteinProphet. Only proteins with minimum ProteinProphet probabilities of 0.9 were considered. Primary interaction data were filtered against a contaminant database resulting in a final list of 242 (replicate A) and 218 (replicate B) bait-prey interactions with an overlap of 85%. To assess the reproducibility rate, the sum of common interactions found in experiments A and B were divided by the total number of interactions identified. **(B)** Generation of a database containing co-purifying contaminant proteins used for data filtering. Eight independent control purifications were performed using SH-tagged eGFP as a bait protein. Tryptic peptides were analysed twice by LC-MS/MS and identified proteins from all 16 measurements were consecutively added to a contaminant database. A total of 109 proteins were identified and used to subtract unspecific protein contaminants. **(C)** Robustness of protein interaction data obtained by the integrated workflow. The number of protein interactions (black line) and the percent overlap (red line) across the two replicate data sets shown as a function of increased filtering stringency (number of unique peptides required for protein identification).

proteins identified from 8 independent SH-eGFP control purifications analysed by 16 LC-MS/MS experiments, which yielded an exhaustive list of 109 contaminant proteins (Figure 4B, Supplementary Table II). Proteins identified in bait-specific experiments that were also present in the contaminant database were considered as unspecific binders and removed from the data set (Supplementary Table V). As ribosomal proteins represent a major group within the contaminant database, we also eliminated all ribosomal proteins as well as keratins introduced during the procedure (for details, see Supplementary Table V). This resulted in a final list of 108 proteins specifically associated with the selected group of bait proteins and resulted in 242 and 218 bait-prey interactions identified in replicate experiments A and B, respectively (Figure 4A, Supplementary Table VI). We found 197 overlapping interactions in the two experiments, which correspond to an overall reproducibility rate of 85%. This is among the highest rates reported so far for systematic AP-MS/MS workflows (Gavin *et al*, 2006).

Applying more stringent filtering criteria by increasing the numbers of unique peptides required for protein identification from 1–4 only marginally increased the overlap from 85 to 92%, but at the same time resulted in a significant decrease (33%) of identified interactions (Figure 4C). To increase the robustness of the final PP2A protein network model, we considered only protein–protein interactions observed in both biological replicate experiments. The high quality of the final network allowed us to identify small but significant differences in the connectivity of closely related proteins. This is illustrated by the data obtained from the two highly connected catalytic subunits PPP2CA and PPP2CB, which share 97% amino-acid sequence identity. All the 28 interacting proteins that we identified in PPP2CA purifications were also found in the group of 32 PPP2CB-associated proteins. The additional four PPP2CB interacting proteins comprise members of the human prefoldin family, indicating that PPP2CB may be linked to the prefoldin complex.

It has been proposed that interactions between two proteins are of higher confidence if the interactions also occur between the corresponding paralogous proteins (Deane *et al*, 2002). Within the 197 interactions identified here, we found 150 interactions for which at least one paralogous interaction was found in our data set (Supplementary Table VII), which serves as an indicator of the achieved data quality and hints at a high degree of combinatorial complex formation by paralogous interactions within the identified phosphatase network. To further validate the data quality, we also compared the interaction data from this study with the existing information from the literature and published databases. Of the 197 interactions identified here, 69 have been reported previously. Among the remaining 128 interactions, we found 30 paralogous and 32 orthologous interactions described previously in humans or yeast, respectively (Figure 5, Supplementary Table VIII). Collectively, these results showed that the method described here provides specific and highly reproducible interaction data sets suited for the generation of robust models of human protein interaction networks.

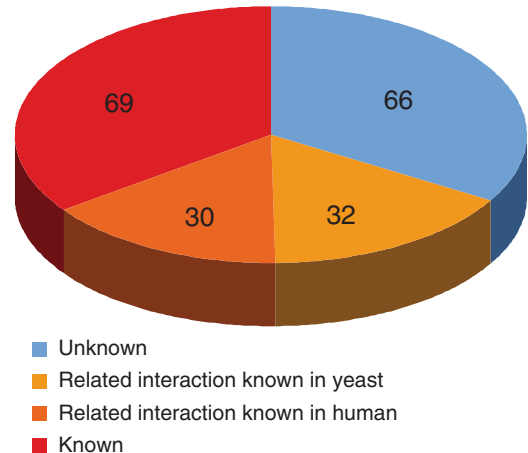


Figure 5 Summary of the identified protein interaction data set. A total of 197 protein–protein interactions identified in this study have been grouped according to available literature information into known interactions, homologous interactions identified in either human or yeast or unknown protein interactions.

A network of human PP2A complexes

The obtained 197 protein interactions are derived from the purification of coexisting phosphatase protein complexes. When the data are assembled into a protein interaction network model, it is, however, difficult to reconstruct the composition of the underlying protein complexes solely from the topology of the resulting network model, unless interaction proteomes have been mapped close to saturation as it has been shown for the yeast proteome (Gavin *et al*, 2006). In the case of the human PP2A phosphatase, however, significant amount of information on the biochemistry and structure of PP2A core complexes have been reported over the past two decades. This information can be used to map the obtained protein interaction data to the corresponding paralogous interactions of known PP2A complexes. This allowed us to assign the protein interactions to groups of related complexes (referred to as modules) and determine which types of the PP2A complexes coexist and can be observed in human embryonic kidney cells by our approach. We found at least five modules of related phosphatase complexes within the obtained network (Figure 6).

PP2A core module

First, we noted a highly connected triangular subnetwork formed between catalytic C, scaffolding A and regulatory B subunits of PP2A. All of the regulatory B subunits were found in association with both the catalytic as well as the scaffolding A subunits of PP2A. No interactions were found between the different catalytic subunits or among the various scaffolding A or the regulatory B subunits. The observed topology is in agreement with literature information and the previously established X-ray structural model on the trimeric PP2A complex (Chao *et al*, 2006; Xu *et al*, 2006; Cho and Xu, 2007). These data therefore suggest the coexistence of a large number of heterotrimeric PP2A core complexes, each consisting of a catalytic, a scaffolding A and a regulatory B subunit (Figure 6, PP2A core module), which represent the majority of

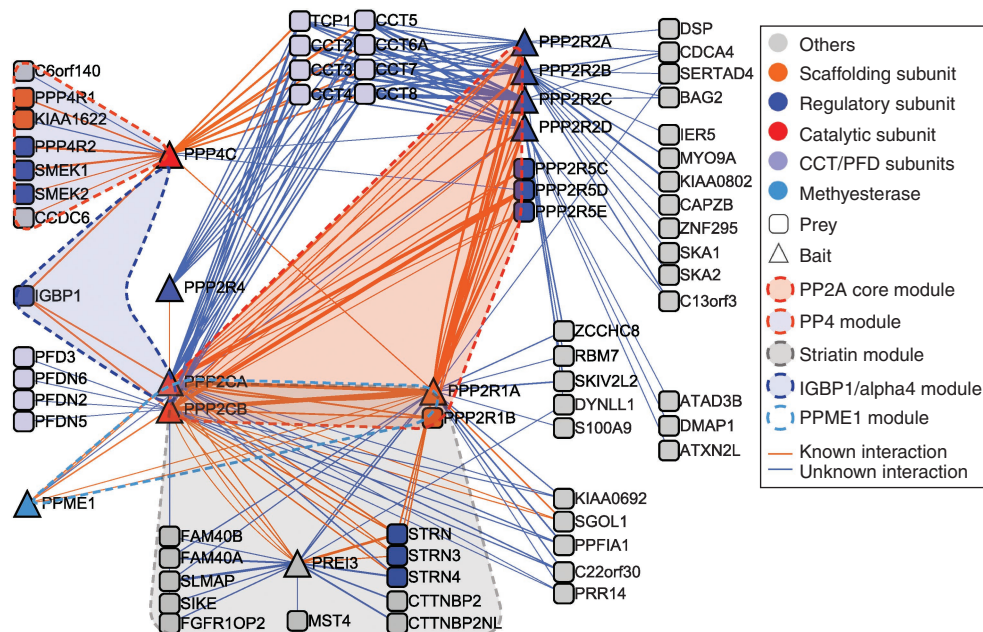


Figure 6 Protein–protein interactions and different classes of protein complexes within the human PP2A interaction network. Protein–protein interaction data identified in two out of two replicate experiments were visualized using Cytoscape v.2.5.1. Bait proteins are represented as triangles, and prey proteins are represented as rounded squares. Nodes were classified according to the legend. Edge colour indicates known interactions (orange) or newly identified interactions (blue). Thickness of the edges reflects the number of identified unique peptides. Five classes of PP2A protein complexes could be distinguished on the basis of network topology and literature information and are displayed by dotted lines and colour shading (see legend).

complexes found in the studied network. These core complexes appear to be connected primarily through the regulatory B subunits to a variety of cellular proteins with functions in cell cycle control, mitosis, transcription and more (see also discussion).

PP4 module

Interestingly, we found regulatory PP2A subunits (PPP2R2D, PPP2R1A and PPP2R5D) also in complexes with PPP4C, the catalytic subunit of the serine/threonine protein phosphatase 4 (PP4). This suggests a certain degree of interchangeable subunit assembly between these two phosphatase systems. When we subsequently used PPP4C as a bait, we also identified proteins exclusively associated with PPP4C representing the second group of phosphatase complexes found in this study (Figure 6, PP4 complexes). These complexes contain the PP4 regulatory proteins PPP4R1 and PPP4R2 as well as proteins linked to DNA damage control, such as SMEK1, SMEK2 and CCDC6 (Gingras *et al*, 2005; Merolla *et al*, 2007), that were absent in the other PP2A complexes analysed. Furthermore, we identified a novel PPP4C-associated protein, KIAA1622, that is characterized by the presence of HEAT and armadillo-like domains. As the same domains are also present in the known phosphatase-scaffolding subunits of PP4 and PP2A, KIAA1622 may represent a novel scaffolding subunit of PP4 complexes.

Striatin module

We were able to identify a highly connected PP2A subnetwork when using PPP2CA, PPP2CB, PPP2R1A and PREI3 (human homologue of yeast Mob1) as bait proteins. Besides a specific

class of regulatory B subunits referred to as the striatins (Strn, Strn3, Strn4), this subnetwork contains the serine/threonine protein kinase MST4, the protein SLMAP as well as specific groups of structurally related proteins (FAM40A, FAM40B; CTTNBP2, CTTNBP2NL; SIKE, FGFR1OP2). The remarkable connectivity among members of this subnetwork provides evidence for a group of higher-order PP2A complexes (Figure 6, striatin module).

IGBP1/alpha4 module

We identified the protein IGBP1 (also referred to as alpha4), the human homologue of yeast TAP42, exclusively in PPP2CA, PPP2CB and PPP4C purifications (Figure 6, IGBP1/alpha4 module). This is in agreement with the present model that IGBP1/alpha4 forms complexes with catalytic subunits in the absence of scaffolding A or regulatory B subunits (Saleh *et al*, 2005). IGBP1/alpha4 was shown to be required for such diverse processes as maintenance of cell survival (Kong *et al*, 2004) as well as learning and memory (Yamashita *et al*, 2006).

PPME1 module

This module is characterized by the presence of the protein phosphatase methylesterase 1 (PPME1), which we have found associated specifically with the catalytic as well as scaffolding subunits of PP2A. PPME1 is crucial for the demethylation of the C-terminal L309 of the PP2A catalytic subunit, which is part of a complex mechanism for the control of PP2A activity (Longin *et al*, 2004; Hombauer *et al*, 2007).

When we mapped our interaction data to the available orthologous interaction data in yeast, we noticed that the interactions within these five modules are highly conserved

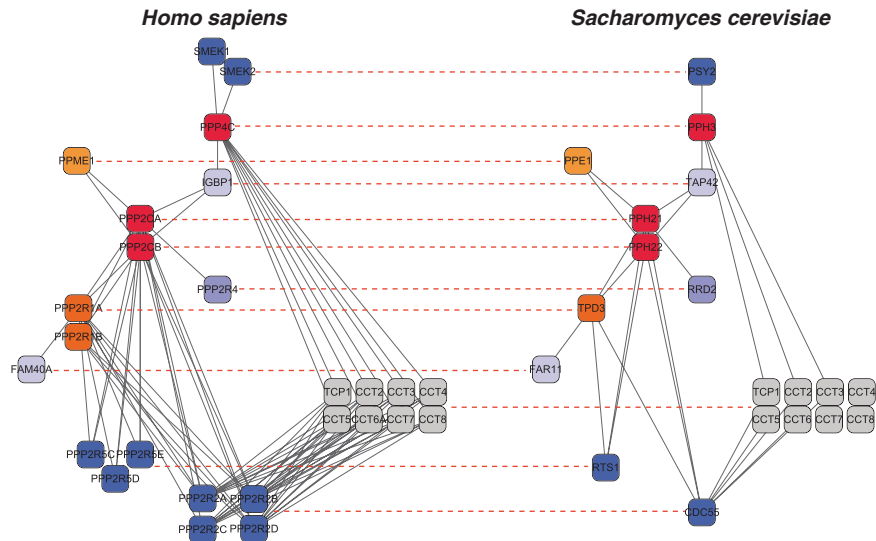


Figure 7 Evolutionary conservation of protein interactions within the PP2A network. Interaction data from this study for which the corresponding orthologous protein interactions could be found in yeast are displayed as a network graph and compared with the yeast PP2A network. Human gene symbols and yeast gene names have been used as node identifiers. Dotted red lines indicate the orthology relationships between the shown network components. Note the expansion of orthologous interactions involving regulatory subunits (dark blue nodes) in the human PP2A system.

during evolution. From this comparison, it became apparent that the number of PP2A regulatory subunits and their interaction with PP2A has expanded in the human system, whereas the number of PP2A catalytic subunits remained constant in human and yeast cells (Figure 7).

In conclusion, the proposed integrated workflow allowed us to assemble the most comprehensive protein network model for human PP2A known to date. The results confirm and significantly extend our knowledge on a remarkable molecular principle underlying the functional diversification of a human phosphatase system by differential protein complex formation.

Discussion

Here, we present an integrated workflow for the systematic analysis of human protein complexes. This workflow, when compared with previous methods, has the following favourable properties: (i) it significantly saves resources, labour and time for the systematic generation of cell lines for inducible expression of affinity-tagged bait proteins; (ii) it has improved sensitivity and thus minimizes the sample amounts required for the successful isolation and analysis of protein complexes; and (iii) it generates protein interaction data of unprecedented reproducibility in human cells suited for data integration in systems biology.

Rapid and reliable production of human cell lines for systematic AP-MS studies

In contrast to studies in yeast where site-specific recombination has facilitated the generation of strains expressing specific affinity-tagged proteins at high throughput, the production of the corresponding human cell lines has been a critical bottleneck for systematic AP-MS studies. Generation of cell

lines in previous studies involved many steps (e.g. PCR cloning, sequencing, selection and testing of cell clones and so on) and consumed significant amounts of resources and time. In addition, the level of bait expression could not be adjusted, which may result in massive protein overexpression of the bait protein.

In the presented method, we used Flp-mediated recombination of specific cDNAs from an orfeome resource for rapid generation of large collections of cell lines. This takes advantage of the rapidly expanding collections of human cDNAs and thus obviates the need for PCR cloning and sequencing. Many of the steps for gateway recombination can be carried out by liquid handling robots and require only a minimum amount of work compared with classical cloning. Starting from an orfeome entry clone, isogenic cell pools for inducible bait expression are reliably obtained within 2–3 weeks, which is significantly less than the time needed with previous methods. In contrast to previous studies, which were almost exclusively performed by industrial labs or large academic consortia (Bouwmeester *et al*, 2004; Ewing *et al*, 2007), our procedure will enable also academic labs with limited personnel resources to conduct high-quality AP-MS studies on a larger scale. Inducible expression as applied in our method significantly expands the range of bait proteins that can be analysed in large-scale AP-MS studies, as cell lines expressing growth-inhibitory and pro-apoptotic bait proteins can now be generated. Importantly, tunable expression of bait proteins from a defined genomic locus as presented here allows equal bait expression at physiological levels throughout the cell population. This reduces the rate of false-positive interactions caused by massive overexpression of bait proteins typically observed in stably or transiently transfected cell lines using constitutive promoters. Collectively, these advances will significantly improve data quality and expand the scope of future large-scale AP-MS studies in human cells.

Increased sensitivity

Another bottleneck in previous large-scale protein complex analysis is linked to the limited overall sensitivities of previous workflows, which require massive amounts of cellular starting material in the range of 1×10^8 – 1×10^9 cells (Al-Hakim *et al*, 2005; Gregan *et al*, 2007). This, in turn, imposes significant problems for high-throughput AP-MS studies with respect to available resources and experimental logistics. Overall sensitivity in AP-MS workflows is determined by both the protein complex purification procedure and the MS analysis step. We introduce a strategy using a streptavidin-binding peptide combined with the hemagglutinin epitope (SH) for the complex purification step in our integrated workflow. The SH-tag consists of 48 amino acids, which is significantly smaller than the 21 kDa TAP tag and allows efficient double-affinity purification with high specificity. Using various bait proteins, we typically obtained overall purification yields between 30 and 40%. These yields are considerably higher than the ones obtained with the classical or the recently improved version of the TAP tag procedure (Burckstummer *et al*, 2006). The significant improvements in purification yield can be attributed to both the efficient binding as well as efficient elution from the two different affinity reagents. The elimination of the TEV cleavage step applied in the classical and modified TAP procedures also accounts for the enhanced yields and overall purification speed (2–3 h). SH-purification yielded complex preparations of high purity suitable for direct LC-MS/MS analysis. This obviates the multiple experimental steps associated with gel-based MS procedures to enable comprehensive protein identification at high throughput from reduced sample amounts. As a result, the combined SH-LC-MS/MS approach allows the identification of specific complex partners from minimal analyte amounts that correspond to 1×10^6 cells. Obviously, more starting material increases the chance to identify also complex components of low abundance. We recommend using around 3×10^7 cells for the proposed procedure, which yields enough material for multiple LC-MS/MS experiments and identification of even substoichiometric components. This amount is about 10 times lower than the amounts used by other workflows (Al-Hakim *et al*, 2005; Burckstummer *et al*, 2006; Gregan *et al*, 2007).

Reproducibility and data robustness

Clearly reliable models on human protein interaction networks at a global scale can only be accomplished as a collective effort by the research community, given the enormous resources involved in this task. This, in turn, will depend on analytical procedures, which minimize the problem of false-positive identification and provide a high level of data robustness.

Co-purifying contaminant proteins are a major source for false-positive interactions. As we show, the SH-purification step in our pipeline allows efficient removal of many of these contaminants. Contaminants that still remained after double-affinity purification were efficiently removed from the

interaction data sets using a contaminant database obtained from multiple SH-eGFP control purifications as a filter.

Data quality in large-scale AP-MS studies largely depends on the reproducibility of the applied experimental strategy. However, very little is known on the overall reproducibility rates of AP-MS workflows at present used for systematic protein complex analysis. To our knowledge, the only reliable source of information was given in a report on the gel-based TAP-MS approach used to study the compendium of yeast complexes, which refers to an overall reproducibility of 69% (Gavin *et al*, 2006). To address this point, we used the presented workflow to analyse a network of human phosphatase complexes in duplicates and found that at least 85% of the identified interaction data overlap between two replicate experiments, which represents a significant advance for future AP-MS studies.

Modularity of a human phosphatase network

Over the past two decades, a considerable amount of biochemical and structural data have been reported on PP2A complexes from different species. These data serve as a valuable resource to determine whether the new high-throughput method can generate validated new biological information. The SH-LC-MS/MS analysis showed the most comprehensive phosphatase protein–protein interaction network model known to date. It consists of 69 protein–protein interactions that were already reported in the published literature and 128 novel interactions, 62 of which were related to previously known interactions (Supplementary Table VIII). The significant overlap with published information again highlights the reliability of the presented analytical AP-MS approach and illustrates its potential to retrieve new biological information. In the case of the studied phosphatase network, this approach showed important insights into the overall modular organization of a human phosphatase system and provided promising clues about the molecular functions of the studied phosphatase complexes.

The network topology obtained by the presented method combined with available literature information showed all major types of PP2A complexes, including the previously described non-canonical PP2A complexes containing alpha4 or the protein phosphatase methyltransferase as well as the canonical heterotrimeric PP2A core complexes that account for the majority of PP2A assemblies we have identified. It is thought that the majority of these canonical PP2A complexes function as heterotrimeric assemblies. Using our analytical method, we could uncover that also higher-order complexes can be formed around these core complexes in human cells. The presented work showed a highly connected PP2A subnetwork, which indicates a novel type of higher-order PP2A complexes in human cells (Figure 6). These complexes are formed by a characteristic set of cellular proteins around a canonical heterotrimeric PP2A core complex that contains a specific group of regulatory B subunits referred to as striatins besides catalytic and scaffolding subunits.

For some of the phosphatase subunits, the nature of associated proteins identified in this study provide important

hints about their potential molecular and cellular functions as demonstrated for the following three examples. (i) PREI3, a central component of novel striatin complexes described above, for example, associates with the serine/threonine kinase MST4, a Ste20-related kinase that mediates cell growth control through the ERK pathway (Lin *et al*, 2001). In this context a recent large-scale AP-MS study has identified the related kinase MST3 associated with proteins of the striatin complex (PREI3, STRN, STRN3), which may point at a novel functional link between PP2A and a subset of Ste20-related kinases (Ewing *et al*, 2007). (ii) In protein phosphatase 4 complexes, we identified SMEK1 and SMEK2 (also referred to as PP4R3 α and PP4R3 β), two proteins with functions in DNA damage control that have been linked previously to PP4 (Gingras *et al*, 2005). In this study, in addition we found CCDC6 as a component of PP4 complexes. CCDC6 has been recently linked to the ATM-dependent DNA damage-response pathway (Merolla *et al*, 2007), which together with the previous results indicates a more general function for PP4 in human DNA damage control. (iii) A subset of PP2A-associated proteins may link PP2A to the control of kinetochore function during mitosis. Among them are SKA1 and SKA2, two novel PP2A-associated proteins that we have identified in complexes with the PP2A regulatory B subunit PPP2R2B. SKA1 and SKA2 are known to bind each other, localize to the kinetochores and have a suggested function in spindle checkpoint silencing (Hanisch *et al*, 2006). Similarly, the known PP2A-associated protein shugoshin, which we identified also in PP2A complexes, is localized also at kinetochores and has an important function in the control of sister chromatid segregation (Salic *et al*, 2004; Kitajima *et al*, 2006). These few examples, which highlight just a subset of the molecular linkages presented in this study, illustrate how the data obtained by the proposed procedure can be used to develop new hypotheses. When compared with the available yeast interaction data, we noted a high degree of overlap between the interactions that constitute the five modules that we found in our network, indicating that these phosphatase modules have evolved relatively early during eukaryotic evolution (Figure 7). Interestingly, it appears that expansion in particular of the regulatory B subunits and their specific interaction with other cellular proteins represents a major step to increase the regulatory opportunities by the PP2A network in multicellular eukaryotes, as the number of PP2A catalytic subunits appear to be the same in humans and yeast.

Systems-oriented concepts on the regulation of biological processes in health and disease are among the major goals in post-genomic biomedical research. Large-scale AP-MS studies will have a central function in this effort, as it is the only method to retrieve global information on how proteins are organized into systems of functionally interacting complexes. This knowledge will be essential to understand the dysfunction of cellular systems in pathological situations and to propose novel routes for interfering with deregulated cellular processes. We are convinced that the advances in throughput, sensitivity and reproducibility collectively achieved by the workflow described here mark a significant step forward towards these goals.

Materials and methods

Generation of expression constructs and tetracycline-inducible cell lines

To generate expression vectors for tetracycline-controlled expression of N-terminally SH-tagged versions of bait proteins, human ORFs provided as pDONR223 vectors were selected from a Gateway-compatible human orfeome library (orfeome v3.1, Open Biosystems) for LR recombination with the in-house-designed destination vector pcDNA5/FRT/TO/SH/GW, which we obtained through ligation of the SH-tag coding sequence and the Gateway recombination cassette into the polylinker of pcDNA5/FRT/TO (Invitrogen).

Flp-In HEK293 cells (Invitrogen) containing a single genomic FRT site and stably expressing the tet repressor were cultured in DMEM (4.5 g/l glucose, 10% FCS, 2 mM L-glutamin, 50 μ g/ml penicillin, 50 μ g/ml streptomycin) containing 100 μ g/ml zeocin and 15 μ g/ml blasticidin. The medium was exchanged with DMEM medium containing 15 μ g/ml blasticidin before transfection. For cell line generation, Flp-In HEK293 cells were co-transfected with the corresponding expression plasmids and the pOG44 vector (Invitrogen) for coexpression of the Flp-recombinase using the Fugene transfection reagent (Roche). Two days after transfection, cells were selected in hygromycin-containing medium (100 μ g/ml) for 2–3 weeks and inducible bait expression was tested by immunoblotting using anti-HA antibodies (Covance). Antibodies for detection of endogenous PPP2R1, PPP2C, STK3 and AMPK α 1 were obtained from Cell Signalling Technologies. Tubulin antibodies were obtained from Sigma, and TIP49 antibodies were prepared as described previously (Gstaiger *et al*, 2003).

Indirect immunofluorescence microscopy

HEK293 cells inducibly expressing SH-tagged bait proteins were seeded on human fibronectin-coated glass coverslips (Becton Dickinson) and grown in DMEM containing 10% FCS. Cells were washed once in PBS before fixation for 15 min in 4% paraformaldehyde/PBS. Fixed cells were washed once with PBS containing 100 mM glycine and twice with PBS. Fixed cells were treated for 2 min with 0.5% Triton/PBS. Following two washes with PBS, cells were incubated with a 1:1,000 dilution of a monoclonal anti-HA monoclonal antibody (HA-11, Covance) for 1 h. Cells were washed three times with PBS and stained with a 1:400 dilution of an anti-mouse-Alexa488 secondary antibody (Invitrogen) and 4',6-diamidino-2-phenylindole (DAPI). Finally, cells were washed three times with PBS and mounted in Vectashield (Vector Laboratories), and images were obtained on a Leica DM5500 Q microscope.

SH-double-affinity purification of protein complexes

Bait expressing Flp-In HEK293 cells were lysed in TNN-HS lysis buffer (0.5% NP40, 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 50 mM NaF, 1.5 mM NaVO₃, 5 mM EDTA, 1 mM DTT, 0.5 mM PMSF and protease inhibitor cocktail (Sigma)). Insoluble material was removed by centrifugation. The cleared lysate was loaded consecutively on spin columns (Bio-Rad) containing 200 μ l Strep-Tactin beads (IBA Technologies) and the beads were washed four times with five bead volumes of TNN-HS lysis buffer. Proteins were eluted from the Strep-Tactin beads with five bead volumes of 2 mM biotin. The eluate was then incubated with 100 μ l anti-HA agarose (Sigma) for 2 h at 4°C on a rotation shaker. Immunoprecipitates were washed four times with 10 bead volumes of TNN-HS lysis buffer and two times with 10 bead volumes of the corresponding lysis buffer in the absence of protease inhibitor and detergent. By adding five bead volumes of 0.2 M glycine, pH 2.5, proteins were eluted and subsequently neutralized with 100 mM NH₄HCO₃. Cysteine bonds were reduced with 5 mM TCEP for 30 min at 37°C and alkylated in 10 mM iodoacetamide at room temperature. Samples were digested by adding 1 μ g trypsin (Promega) overnight at 37°C. The peptides were purified using C18 microspin columns (Harvard Apparatus) according to the protocol of the manufacturer,

resolved in 0.1% formic acid and 1% acetonitrile and injected into the mass spectrometer.

LC-MS/MS analysis

LC-MS/MS analysis was performed using an Agilent 1100 series pump (Agilent Technologies, Waldbronn, Germany) and an LTQ mass spectrometer (Thermo Electron, Bremen, Germany). The setup of the μ RPLC system and the capillary column were described previously (Yi *et al*, 2003). The electrospray voltage was set to 1.8 kV. The samples were loaded from the Agilent auto sampler onto the precolumn at a flow rate of 4 μ l/min in 5 min. Mobile phase A was 0.1% formic acid and mobile phase B was 100% acetonitrile (Sigma, Buchs, Switzerland). For analysis, a separating gradient from 5 to 45% mobile phase B over 40 min at 0.2 μ l/min was applied. The three most abundant precursor ions in each MS scan were selected for CID if the intensity of the precursor ion exceeded 20 000 ion counts. Dynamic exclusion window was set to 2 min. To prevent cross-contamination between the different samples, the LC system was washed with a pulse of 8 μ l trifluoroethanol (Fluka, Buchs, Switzerland) after every sample. A peptide standard containing 200 fmol of [Glu1]-Fibrinopeptide B human (Sigma, Buchs, Switzerland) was analysed by LC-MS/MS to constantly monitor the performance of the LC-MS/MS system. Samples that were subjected to label free quantification were analysed on a high performance LTQ-FT-ICR mass spectrometer (Thermo Electron, Bremen, Germany), which was connected online to a nanoelectrospray ion source (Thermo Electron, Bremen, Germany). Peptide separation was carried out using an Eksigent Tempo Nano LC System (Eksigent Technologies, Dublin, CA, USA) equipped with an RP-HPLC column (75 μ m \times 15 cm) packed in-house with C18 resin (Magic C18 AQ 3 μ m; Michrom BioResources, Auburn, CA, USA) using a linear gradient from 96% solvent A (0.15% formic acid, 2% acetonitrile) and 4% solvent B (98% acetonitrile, 0.15% formic acid) to 25% solvent B over 60 min at a flow rate of 0.3 μ l/min. The data acquisition mode was set to obtain one high-resolution MS scan in the ICR cell at a resolution of 100 000 full-width at half maximum (at m/z 400) followed by MS/MS scans in the linear ion trap of the three most intense ions (overall cycle time of 1 s). To increase the efficiency of MS/MS attempts, the charged state screening modus was enabled to exclude unassigned and singly charged ions. Only MS precursors that exceeded a threshold of 150 ion counts were allowed to trigger MS/MS scans. The ion accumulation time was set to 500 ms (MS) and 250 ms (MS/MS) using a target setting of 10^6 (for MS) and 10^4 (for MS/MS) ions. After every sample, a peptide mixture containing 200 fmol of [Glu1]-Fibrinopeptide B human (Sigma, Buchs, Switzerland) was analysed by LC-MS/MS to constantly monitor the performance of the LC-MS/MS system. All raw data have been made publicly accessible through the PeptideAtlas database (<http://www.peptideatlas.org/repository/publications/Glatter2008/>).

MS2 peptide assignments and MS1 alignment

Acquired MS2 scans were searched against the human International Protein Index (IPI) protein database (v.3.26) using the X-Tandem search algorithm (Craig and Beavis, 2004) with k-score plug-in (MacLean *et al*, 2006). *In silico* trypsin digestion was performed after lysine and arginine (unless followed by proline) in fully tryptic peptides. Allowed monoisotopic mass error for the precursor ions was 3 Da for the LTQ data and 50 p.p.m. for the FT data. A fixed residue modification parameter was set for carboxyamidomethylation (+ 57.021464 Da) of cysteine residues. Oxidation of methionine (+ 15.994915 Da) was set as variable residue modification parameter. Model refinement parameters were set to allow phosphorylation (+ 79.966331 Da) of serine, threonine and tyrosine residues as variable modifications. Furthermore, semi-tryptic peptides were allowed for refinement searches. For scoring, a maximum of two missed cleavages were considered. Search results were evaluated on the Trans Proteomic Pipeline (TPP v3.2) using PeptideProphet (Keller *et al*, 2002) and ProteinProphet (Nesvizhskii *et al*, 2003; Keller *et al*, 2005). For label-free quantification, samples were analysed on an LTQ-FT-ICR instrument and MS1 alignment was performed using Superhirn (Mueller *et al*, 2007; Rinner *et al*, 2007). Protein intensity profiles were normalized using the

average $\log_2 (I_i^{(1,2 \text{ or } 3)}) / (I_i^{(1)})$ ($I_i^{(j)}$ = ion intensity of feature i in dilution j) over features belonging to the same protein and normalized to the bait protein profile.

Analysis of protein interaction data

Protein identifications with ProteinProphet probabilities < 1 were subjected to manual inspection to exclude misassigned protein identifications. Protein IDs were filtered against a contaminant database obtained from a total of eight independent SH-eGFP control purifications analysed by 16 LC-MS/MS experiments and mapped to non-redundant entrez gene IDs. Interaction data from two independent replicates were compared and interactions found in both replicates were used for assembly of protein-protein interaction network models using Cytoscape 2.5.1 (www.cytoscape.org). All interaction data have been made publicly accessible through the IntAct protein interaction database (<http://www.ebi.ac.uk/intact/site/index.jsf>). To identify known interactions, an in-house database containing 61 263 interactions from various databases (BIND, DIP, IntAct, HPRD) and published literature (Ramani *et al*, 2005; Rual *et al*, 2005; Stelzl *et al*, 2005; Ewing *et al*, 2007) was used. Orthology data have been retrieved from the Ensemble database (version from 16 Sep 2008) using biomart version 0.7 (<http://www.biomart.org>). Interaction data from yeast have been downloaded from the biogrid server <http://www.thebiogrid.org> (Stark *et al*, 2006) and Cytoscape 2.5.1 was used for mapping orthologous interactions.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Hemmo Meyer and Martin Beck for helpful discussions and Oliver Rinner for critically reading the manuscript. This project has been funded in part by the ETH Zurich, and with Federal funds from the National Heart, Lung and Blood Institute, National Institutes of Health, under contract no. N01-HV-28179. AW and RA were supported in part by a grant from F Hoffmann-La Roche Ltd (Basel, Switzerland) provided to the Competence Center for Systems Physiology and Metabolic Disease. TG was supported by a TH grant of the ETH Zürich and a fellowship from the Roche Research Foundation.

Author contributions

TG conducted most of work to establish the procedure. AW and TG performed purification experiments and data analysis. RA provided mass spectrometry concepts and contributed to writing of the manuscript. MG planned and supervised the project and wrote most of the manuscript. All authors discussed the results and commented on the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Al-Hakim AK, Goransson O, Deak M, Toth R, Campbell DG, Morrice NA, Prescott AR, Alessi DR (2005) 14-3-3 cooperates with LKB1 to regulate the activity and localization of QSK and SIK. *J Cell Sci* **118**: 5661–5673
- Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, Cruciat C, Eberhard D, Gagneur J, Ghidelli S, Hopf C, Huhse B, Mangano R, Michon AM, Schirle M, Schlegl J, Schwab M,

- Stein MA, Bauer A, Casari G *et al* (2004) A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nat Cell Biol* **6**: 97–105
- Burckstummer T, Bennett KL, Preradovic A, Schutze G, Hantschel O, Superti-Furga G, Bauch A (2006) An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods* **3**: 1013–1019
- Chao Y, Xing Y, Chen Y, Xu Y, Lin Z, Li Z, Jeffrey PD, Stock JB, Shi Y (2006) Structure and mechanism of the phosphotyrosyl phosphatase activator. *Mol Cell* **23**: 535–546
- Cho US, Xu W (2007) Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445**: 53–57
- Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**: 1466–1467
- Deane CM, Salwinski L, Xenarios I, Eisenberg D (2002) Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* **1**: 349–356
- Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y *et al* (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**: 89
- Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffmann V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M *et al* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**: 631–636
- Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T *et al* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**: 141–147
- Gingras AC, Caballero M, Zarske M, Sanchez A, Hazbun TR, Fields S, Sonenberg N, Hafen E, Raught B, Aebersold R (2005) A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol Cell Proteomics* **4**: 1725–1740
- Gingras AC, Gstaiger M, Raught B, Aebersold R (2007) Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**: 645–654
- Gravel RA, Narang MA (2005) Molecular genetics of biotin metabolism: old vitamin, new science. *J Nutr Biochem* **16**: 428–431
- Gregan J, Riedel CG, Petronczki M, Cipak L, Rumpf C, Poser I, Buchholz F, Mechtler K, Nasmyth K (2007) Tandem affinity purification of functional TAP-tagged proteins from human cells. *Nat Protoc* **2**: 1145–1151
- Gstaiger M, Luke B, Hess D, Oakeley EJ, Wirbelauer C, Blondel M, Vigneron M, Peter M, Krek W (2003) Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. *Science* **302**: 1208–1212
- Hanisch A, Sillje HH, Nigg EA (2006) Timely anaphase onset requires a novel spindle and kinetochore complex comprising Ska1 and Ska2. *EMBO J* **25**: 5504–5515
- Hart GT, Ramani AK, Marcotte EM (2006) How complete are current yeast and human protein-interaction networks? *Genome Biol* **7**: 120
- Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreau M, Muskat B, Alfarano C *et al* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**: 180–183
- Hombauer H, Weismann D, Mudrak I, Stanzel C, Fellner T, Lackner DH, Ogris E (2007) Generation of active protein phosphatase 2A is coupled to holoenzyme assembly. *PLoS Biol* **5**: e155
- Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* **98**: 4569–4574
- Janssens V, Goris J (2001) Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochem J* **353**: 417–439
- Janssens V, Goris J, Van Hoof C (2005) PP2A: the expected tumor suppressor. *Curr Opin Genet Dev* **15**: 34–41
- Junttila MR, Saarinen S, Schmidt T, Kast J, Westermarck J (2005) Single-step Strep-tag purification for the isolation and identification of protein complexes from mammalian cells. *Proteomics* **5**: 1199–1203
- Keller A, Eng J, Zhang N, Li XJ, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **1**: 2005 0017
- Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**: 5383–5392
- Kitajima TS, Sakuno T, Ishiguro K, Iemura S, Natsume T, Kawashima SA, Watanabe Y (2006) Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441**: 46–52
- Kocher T, Superti-Furga G (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* **4**: 807–815
- Kong M, Fox CJ, Mu J, Solt L, Xu A, Cinali RM, Birnbaum MJ, Lindsten T, Thompson CB (2004) The PP2A-associated protein alpha4 is an essential inhibitor of apoptosis. *Science* **306**: 695–698
- Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B *et al* (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**: 637–643
- Lamesch P, Li N, Milstein S, Fan C, Hao T, Szabo G, Hu Z, Venkatesan K, Bethel G, Martin P, Rogers J, Lawlor S, McLaren S, Dricot A, Borick H, Cusick ME, Vandenhaute J, Dunham I, Hill DE, Vidal M (2007) hORFeome v3.1: a resource of human open reading frames representing over 10 000 human genes. *Genomics* **89**: 307–315
- Lechward K, Awotunde OS, Swiatek W, Muszynska G (2001) Protein phosphatase 2A: variety of forms and diversity of functions. *Acta Biochim Pol* **48**: 921–933
- Leulliot N, Vicentini G, Jordens J, Quevillon-Cheruel S, Schiltz M, Barford D, van Tilbeurgh H, Goris J (2006) Crystal structure of the PP2A phosphatase activator: implications for its PP2A-specific PPlase activity. *Mol Cell* **23**: 413–424
- Lin JL, Chen HC, Fang HI, Robinson D, Kung HJ, Shih HM (2001) MST4, a new Ste20-related kinase that mediates cell growth and transformation via modulating ERK pathway. *Oncogene* **20**: 6559–6569
- Liu H, Sadygov RG, Yates III JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**: 4193–4201
- Longin S, Jordens J, Martens E, Stevens I, Janssens V, Rondelez E, De Baere I, Derua R, Waelkens E, Goris J, Van Hoof C (2004) An inactive protein phosphatase 2A population is associated with methylesterase and can be re-activated by the phosphotyrosyl phosphatase activator. *Biochem J* **380**: 111–119
- MacLean B, Eng JK, Beavis RC, McIntosh M (2006) General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics* **22**: 2830–2832
- Merolla F, Pentimalli F, Pacelli R, Vecchio G, Fusco A, Grieco M, Celetti A (2007) Involvement of H4(D10S170) protein in ATM-dependent response to DNA damage. *Oncogene* **26**: 6167–6175
- Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R, Muller M (2007) SuperHirn—a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **7**: 3470–3480
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**: 4646–4658
- O'Gorman S, Fox DT, Wahl GM (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**: 1351–1355

- Ramani AK, Bunescu RC, Mooney RJ, Marcotte EM (2005) Consolidating the set of known human protein–protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol* **6**: R40
- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**: 1030–1032
- Rinner O, Mueller LN, Hubalek M, Muller M, Gstaiger M, Aebersold R (2007) An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nat Biotechnol* **25**: 345–352
- Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S *et al* (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**: 1173–1178
- Saleh M, Pagliari LJ, Green DR (2005) Life support: the alpha4 phosphatase subunit in cell survival and apoptosis. *Trends Cell Biol* **15**: 285–287
- Salic A, Waters JC, Mitchison TJ (2004) Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. *Cell* **118**: 567–578
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**: D535–D539
- Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlauff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B *et al* (2005) A human protein–protein interaction network: a resource for annotating the proteome. *Cell* **122**: 957–968
- Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**: 623–627
- Virshup DM (2000) Protein phosphatase 2A: a panoply of enzymes. *Curr Opin Cell Biol* **12**: 180–185
- Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**: 116–122
- Westermarck J, Hahn WC (2008) Multiple pathways regulated by the tumor suppressor PP2A in transformation. *Trends Mol Med* **14**: 152–160
- Xing Y, Li Z, Chen Y, Stock JB, Jeffrey PD, Shi Y (2008) Structural mechanism of demethylation and inactivation of protein phosphatase 2A. *Cell* **133**: 154–163
- Xu Y, Xing Y, Chen Y, Chao Y, Lin Z, Fan E, Yu JW, Strack S, Jeffrey PD, Shi Y (2006) Structure of the protein phosphatase 2A holoenzyme. *Cell* **127**: 1239–1251
- Yamashita T, Inui S, Maeda K, Hua DR, Takagi K, Fukunaga K, Sakaguchi N (2006) Regulation of CaMKII by alpha4/PP2Ac contributes to learning and memory. *Brain Res* **1082**: 1–10
- Yi EC, Lee H, Aebersold R, Goodlett DR (2003) A microcapillary trap cartridge–microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation. *Rapid Commun Mass Spectrom* **17**: 2093–2098



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Licence.