nature biotechnology

An integrated mass spectrometric and computational framework for the analysis of protein interaction networks

Oliver Rinner[1,4], Lukas N. Mueller[1,4], Martin Hubálek[2], Markus Müller[1], Matthias Gstaiger[1] & Ruedi Aebersold[1,3]

Biological systems are controlled by protein complexes that associate into dynamic protein interaction networks. We describe a strategy that analyzes protein complexes through the integration of label-free, quantitative mass spectrometry and computational analysis. By evaluating peptide intensity profiles throughout the sequential dilution of samples, the MasterMap system identifies specific interaction partners, detects changes in the composition of protein complexes and reveals variations in the phosphorylation states of components of protein complexes. We use the complexes containing the human forkhead transcription factor FoxO3A to demonstrate the validity and performance of this technology. Our analysis identifies previously known and unknown interactions of FoxO3A with 14-3-3 proteins, in addition to identifying FoxO3A phosphorylation sites and detecting reduced 14-3-3 binding following inhibition of phosphoinositide-3 kinase. By improving specificity and sensitivity of interaction networks, assessing post-translational modifications and providing dynamic interaction profiles, the MasterMap system addresses several limitations of current approaches for protein complexes.

Most biological processes are controlled by dynamic molecular networks of enormous complexity. Several methods for the analysis of protein complexes and protein interactions have been developed, and some have been applied in large-scale studies[1–8]. Among these, analysis of affinity-purified protein complexes by liquid chromatography mass spectroscopy (LC-MS) has been particularly attractive, because, in principle, all the components even of large complexes can be identified in a single experiment[9]. However, these methods have at least three limitations. First, the false-positive error rates are generally large and can be reliably estimated only at the level of a whole data set in large-scale protein interaction experiments, where the reproducibility is typically only ~70% (ref. 3). Second, although their structures and compositions change dynamically as a function of cellular conditions, protein complexes and networks are represented as static entities. Third, usually no information about the state of post-translational modifications is collected, even though they frequently modulate the compositions and the subcellular locations of complexes. Quantitative MS approaches based on isotope labeling[10–14], along with appropriate control experiments, can distinguish background contaminants in protein complex purifications from true interactors[15–17] and identify changes in the compositions of protein complexes[18]. Specifically, *in vivo* labeling is expensive and cannot be applied for the analysis of tissue samples from animals or patients. Furthermore, the quantification of peptide ratios is limited to peptides carrying the isotope tag and depends on their successful identification by tandem mass spectroscopy (MS/MS). An alternative approach is to use label-free quantification of peptide signals (features), which are identified in different samples based on their MS1 ion currents[19–21]. For instance, peptide profiles from centrifugation fractions were used in combination with localization information to identify components of the centrosome[22].

We present an integrated computational and mass spectrometric strategy for the analysis of protein complexes from co-immunoprecipitations (Co-IP), which combines the generation of highly accurate LC-MS patterns using a linear ion-trap Fourier transform mass spectrometer and algorithms to detect and map features across different LC-MS runs. This information is then unified in a MasterMap, from which MS1 features with quantitative or qualitative differences are selected for subsequent targeted MS2 experiments. We use the MasterMap concept to address three major technical issues in the systematic analysis of protein interaction networks: increase of specificity and sensitivity, quantification of protein interaction dynamics and identification of sites of protein phosphorylation.

RESULTS

A strategy for the dissection of protein-protein interactions

We chose as a model human complexes containing FoxO3A, a member of the forkhead domain transcription factor family that is regulated by signals such as insulin, stress response and nutrient availability[23]. Responsiveness to any of these signals is mediated by

[1]Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. [2]Institute of Molecular Pathology, Faculty of Military Health Sciences, University of Defence, Trebesska 1575, Hradec Králové, Czech Republic. [3]Faculty of Science, University of Zurich, Zurich, Switzerland. [4]These authors contributed equally to this work. Correspondence should be addressed to R.A. (aebersold@imsb.biol.ethz.ch) or M.G. (matthias.gstaiger@imsb.biol.ethz.ch) or M.M. (markus.mueller@imsb.biol.ethz.ch).

Received 21 December 2006; accepted 18 January 2007; published online 25 February 2007; doi:10.1038/nbt1289

post-translational modifications of FoxO3A and its association with other proteins.

To study FoxO3A protein interactions, we performed standard Co-IP experiments from cell extracts using hemagglutinin (HA) epitope–tagged FoxO3A as the bait protein. MS1 signals from several LC-MS experiments were integrated into a MasterMap (**Supplementary Results** online). To identify specific interaction partners of FoxO3A, we compared the bait sample to a Co-IP sample prepared in parallel with extracts from a control cell line (**Fig. 1a**). Relative peptide amounts were quantified based on the linear properties of the mass spectrometer over a wide dynamic range (**Supplementary Fig. 1** online). Intensity ratios fail to separate bait-specific interaction partners from background samples because ratios cannot be interpreted when a signal is detected in only one sample (**Supplementary Fig. 2** online). To circumvent this problem the bait sample was sequentially diluted into the control sample. Consequently, FoxO3A and its specific interaction partners were progressively enriched, whereas the concentrations of contaminant proteins present in both samples remained constant in all four dilution steps.

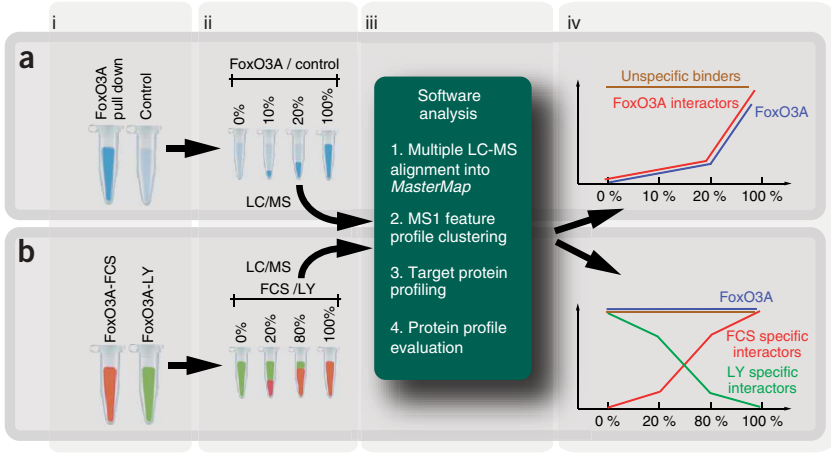

Figure 1 Schematic overview of the experimental approach. (**a**) Peptides are prepared by co-immunoprecipitation from a FoxO3A-expressing cell line and from a control cell line (i). The samples are mixed in a four-step dilution series (ii), and then analyzed by LC-MS and subjected to quantification by the software SuperHirn (iii). Extracted protein profiles allow the categorization of unspecific binding proteins (brown line) and FoxO3A interaction partners (blue line) (iv). (**b**) To quantify dynamic changes in the FoxO3A interaction pattern, a dilution series between Co-IP experiments performed using cells grown in the presence of FCS or subjected to serum starvation and exposure to the PI3K inhibitor Ly294002 (LY condition) was prepared (i,ii). Samples were subjected to LC-MS and analyzed by SuperHirn (iii). Subsequently, peptide and profiles were analyzed to quantify changes in the abundance profile of previously identified interaction partners of FoxO3A (iv).

Each dilution sample was analyzed by LC-MS and the SuperHirn software developed in-house was used to extract protein profiles by MS1 feature profiling and classify them into profile groups by *k*-means clustering analysis. **Figure 1a** (panel iv) illustrates the expected abundance profiles of interaction partners and contaminant proteins. Whereas the abundance of FoxO3A (red line) and its interaction partners (blue line) increases, the levels of contaminants (brown line) remain constant. As discussed below, essentially the same process can be used to detect changes in the composition of protein complexes associated with different cellular states (**Fig. 1b**).

Interaction of FoxO3A with members of the 14-3-3 family

As expected, the profile group closest to the theoretical dilution profile contained the peptides from the FoxO3A bait protein (**Supplementary Fig. 3** online). This cluster defines the target cluster and supposedly contains peptides from proteins that purify specifically with the bait. Peptides were assembled into proteins based on high confidence MS2 identifications (PeptideProphet peptide probability[24] > 0.9) and protein profiles were constructed. The protein profiles are clearly distinguishable as either enriched or constant, thereby facilitating the identification of potential binding partners of FoxO3A through the strong correlation of their profiles with the target cluster profile (**Fig. 2a**). Accordingly, a histogram of the profile scores reveals two populations of contaminates and potential specific interactors, which were modeled by Gaussian mixtures for statistical evaluation of the correlation between the protein profiles and the target cluster profile (**Fig. 2b**).

Table 1 lists the proteins that were identified by a high protein profile correlation to the target cluster profile (profile probability > 0.9) and therefore represent likely FoxO3A interaction partners. FoxO3A is distinctly identified by numerous peptide hits and six different

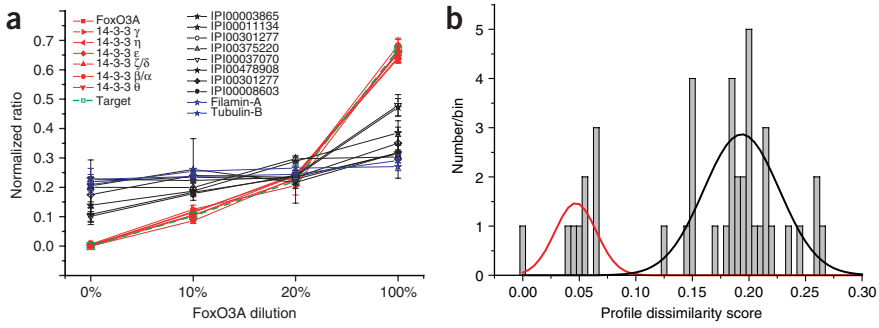

Figure 2 Enrichment profiles of protein groups separate specific interaction partners from contaminant proteins. (**a**) Normalized protein profiles in dilutions with peptides from a control Co-IP (FoxO3A-FCS). The green line indicates the profile of the target cluster derived from unsupervised clustering, which contains the FoxO3A bait protein. Profiles in red refer to proteins of the 14-3-3 family with high correlation to the target cluster profile. Black lines show the profiles of the six next best proteins with >1 peptide member(s). These proteins are only slightly enriched and clearly present in the control sample. Blue profiles represent typical contaminant proteins that were identified with many peptides. Error bars indicate s.d. of all peptides assigned to a protein. (**b**) The cluster with the highest correlation to the theoretical dilution schema was selected as the target cluster. Protein profiles were subsequently extracted from the three closest clusters and scored against the target cluster profile. Profile correlation scores were separated by expectation maximization of a two-parameter Gaussian mixture model. Small scores indicate high similarity to the target cluster profile.

Table 1 Identified interaction partners of FoxO3A

			Number of members *M*				Profile evaluation	
	IPI identifier	Protein description	M^a	M^b	M^c	M^d	Score	Probability
FCS	IPI00012856	FoxO3A	27	13	+1	+11	$4.42 \cdot 10^{-3}$	1.00000
	IPI00216318	14-3-3 β/α	5	1	+1	+5	$5.09 \cdot 10^{-3}$	1.00000
	IPI00000816	14-3-3 ε	13	5	+2	+7	$7.78 \cdot 10^{-3}$	1.00000
	IPI00018146	14-3-3 τ	4	4	–	+4	$9.48 \cdot 10^{-3}$	1.00000
	IPI00216319	14-3-3 η	6	4	–	+2	$1.07 \cdot 10^{-2}$	1.00000
	IPI00021263	14-3-3 ζ/δ	8	1	+3	+4	$1.25 \cdot 10^{-2}$	0.99995
	IPI00220642	14-3-3 γ	5	3	–	+2	$2.16 \cdot 10^{-2}$	0.99924
LY	IPI00220642	14-3-3 γ	1	1	–	+1	$6.98 \cdot 10^{-2}$	1.00000
	IPI00012856	FoxO3A	15	10	+2	+5	$7.43 \cdot 10^{-2}$	0.99907
	IPI00021263	14-3-3 ζ/δ	2	2	+1	+2	$8.56 \cdot 10^{-2}$	0.99690

Protein profiles were scored according to similarity to the target cluster profile. Assigned profile probability values reflect the likelihood for a true similarity with the target cluster profile. A cut-off of $P > 0.9$ was used as significance threshold. Proteins without unique peptides were removed.
[a]Number of MS1 features. [b]Number of identified MS1 features. [c]Increase in MS1 features by inclusion list and LTQ data.
[d]Increase in identified peptides by PMM.

subforms (β/α, γ, ε, η, τ, ζ/δ) of the 14-3-3 protein family are listed as FoxO3A interaction partners on the basis of their high profile probability. This observation confirms the generally accepted notion that 14-3-3 proteins regulate FoxO3A activity[25] and extends the list of 14-3-3 proteins that bind to FoxO3A. Only 14-3-3 ζ/δ was previously reported to bind FoxO3A[26]. These results were confirmed in a replicate experiment with post-lysis cross-linking (see **Supplementary Results** and **Supplementary Fig. 4** online).

Quantification of growth state–dependent protein-interactions

The protein-interaction pattern described above was derived under growth-promoting conditions in the presence of fetal calf serum (FCS condition). To reveal growth state–specific changes in the interaction pattern of FoxO3A, we repeated the experiment for cells subjected to growth inhibition by serum starvation plus inhibition of PI3K with the drug Ly294002 (ref. 27) (LY condition). Immunofluorescence microscopy confirmed that our cell line responded to PI3K inhibition by substantial translocation of FoxO3A into the nucleus as described previously[26] (**Fig. 3a**).

Compared to the growth-promoting condition, fewer members of the 14-3-3 family were identified as binding partners (**Table 1**). Of the previously identified six 14-3-3 members, only the proteins 14-3-3 γ and ζ/δ were unambiguously detected as FoxO3A interaction partners when PI3K was inhibited. The reduced appearance of 14-3-3 interactions with FoxO3A following PI3K inactivation is consistent with previous results obtained using Co-IP western blotting[27].

Although it is tempting to infer changes in protein-protein interactions from the comparison of the list of peptides identified by MS/MS, these data may not correctly indicate the true quantitative nature of the underlying interaction dynamics. Sampling of precursor ions for fragmentation has low reproducibility and depends on both the amount and complexity of the sample, as well as the ion intensity distributions (**Supplementary Fig. 5** online). We therefore quantified changes in the identified FoxO3A interaction partners between the FCS and LY growth conditions at the MS1 level. Again, samples of FoxO3A-FCS and FoxO3A-LY were mixed to obtain profiles that could be checked for consistency. As binding partners could possibly show an increasing or decreasing profile, depending on their response to growth factor signaling, we used a symmetrical mixing scheme (**Fig. 1b**). As profiles of proteins present in both samples (e.g., FoxO3A itself, contaminants and interaction partners insensitive to signaling changes) are flat, differentially enriched proteins can be easily discriminated owing to the effects of sample dilution on the degree of enrichment (**Fig. 1d**).

Levels of all 14-3-3 proteins exhibit more than a twofold reduction under growth-inhibiting conditions (**Fig. 3b** and **Table 2**). Interaction of FoxO3A with 14-3-3 proteins was not completely repressed in the LY condition, confirming that FoxO3A does not completely translocate to the nucleus in all Ly294002-treated cells[26] or that nuclear FoxO3A does not completely lose its binding to 14-3-3 (ref. 27). These data also indicate that quantitative analysis of the MS1 patterns results in a substantially more complete and accurate picture of the FoxO3A complex dynamics than the MS2 data alone.

MS1 feature profiling for targeted peptide annotation

The MasterMap is based on the alignment of MS1 features detected in multiple LC-MS runs. High confidence MS2 information acquired during the LC-MS runs is used in the final data-processing steps where peptides are assembled into proteins. Therefore, quantification of peptide signals at the MS1 level enables the classification of features

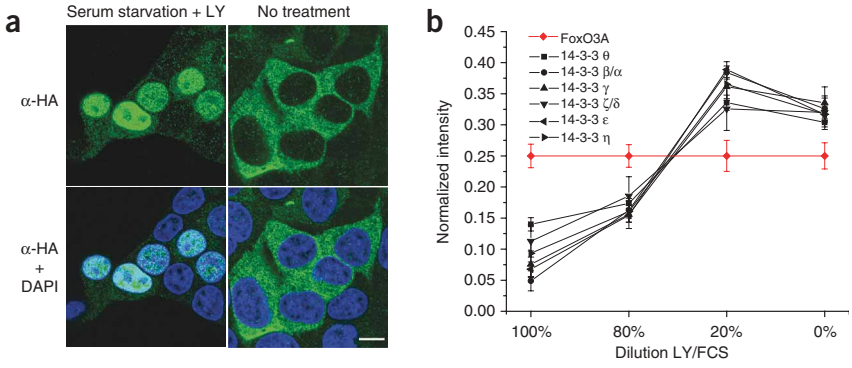

Figure 3 The MasterMap reveals quantitative changes in the abundance of FoxO3A interaction partners upon PI3K inhibition. (**a**) Nuclear translocation of HA-FoxO3A caused by serum deprivation and PI3K inhibition. Cells stably expressing HA-FoxO3A were either serum starved overnight and then treated with 20 μM LY294002 inhibitor (LY) or grown in the presence of 10% FCS (no treatment) and analyzed by laser confocal immunofluorescence microscopy using an anti HA antibody (12CA5) and DAPI. Scale bar, 10 μm. (**b**)14-3-3 protein association is reduced under growth-inhibiting conditions. Peptide samples from Co-IPs carried out under growth-promoting and PI3K-inhibiting conditions were mixed in four steps (**Fig. 1**). Protein profiles were normalized against FoxO3A and were translated into enrichment factors based on similarity with a theoretical dilution model. Error bars indicate s.e.m.

Table 2 Growth state–specific changes in the FoxO3A interaction pattern

		Enrichment factor	
		TNN	DSP
IPI identifier	Protein description	LY/FCS	LY/FCS
IPI00012856	FoxO3A	1.0 ± 0.23	1.00 ± 0.16
IPI00216318	14-3-3 β/α	0.19 ± 0.05	0.55 ± 0.12
IPI00021263	14-3-3 ζ/δ	0.46 ± 0.19	0.53 ± 0.08
IPI00220642	14-3-3 γ	0.39 ± 0.10	0.63 ± 0.13
IPI00216319	14-3-3 η	0.41 ± 0.17	0.43 ± 0.09
IPI00000816	14-3-3 ε	0.40 ± 0.15	0.55 ± 0.11
IPI00018146	14-3-3 τ	0.26 ± 0.05	0.69 ± 0.10
IPI00554737	PP2A 65 kDa, subunit Aα	–	2.04 ± 0.28
IPI00008380	PP2A catalytic subunit α	–	1.73 ± 0.16
IPI00556528	PP2A 56 kDa, subunit Bε		1.33 ± 0.07

A dilution mix was created from Co-IP samples collected under LY and FCS conditions and protein profiles of previously identified FoxO3A interactors were extracted. Dilution factor–corrected ratios were calculated between the two CoIP experiments LY and FCS normalized to the abundance of the bait FoxO3A. The experiment was repeated for samples treated with DSP cross linker (**Supplementary Results**). Error values indicate s.e.m. All reported ratios are significantly different from 1 (z-test, one-sided, $P < 1\%$).

into enriched peptides or unspecific binders, even in the absence of MS2 information. This unique property of the MasterMap was used to increase peptide identification coverage and to detect post-translationally modified peptide variants. MS1 features encoded in the MasterMap are categorized by profile clustering into specific or contaminant proteins and peptide annotation efforts are focused on the list of MS1 features displaying a high correlation to the target cluster profile (**Fig. 4a**). In addition to the conventional integration of MS2 scans from the performed LC-MS runs, three strategies were used to annotate the list of enriched MS1 features: targeted peptide sequencing using mass to charge (*m*/*z*) inclusion lists, peptide mass mapping (PMM) and integration of MS2 identifications from LC-MS/MS data sets acquired using other mass spectrometers.

In the inclusion list approach, the peptide identities of the selected MS1 features were assessed by targeted MS2 sequencing in additional LC-MS runs. In parallel, we measured aliquots of the samples on a Thermo linear ion trap (LTQ) mass spectrometer, which is optimized for acquiring large numbers of MS2 scans at high sensitivity. Obtained high confidence MS2 identifications (PeptideProphet peptide probability > 0.9) were assigned to MS1 features stored in the MasterMap using their theoretically calculated molecular peptide mass.

Overall, the extended MS1 feature annotation approach increased the number of assigned MS1 features per protein, thus improving the protein sequence coverage (**Table 1** and **Supplementary Table 1** online). Moreover, the detection of additional proteotypic peptides increased the confidence with which specific protein isoforms, for example, 14-3-3 family members, could be distinguished.

Figure 4b shows the effect of the additional identifications of clustered MS1 features exemplified for the 14-3-3 family in a single experiment. It also illustrates the challenge of inferring the presence of proteins by conventional MS2 information only[28,29]. Peptides that connect to two or more proteins are nonproteotypic peptides, which do not unambiguously identify a single protein. Whereas 14-3-3 ζ/δ,

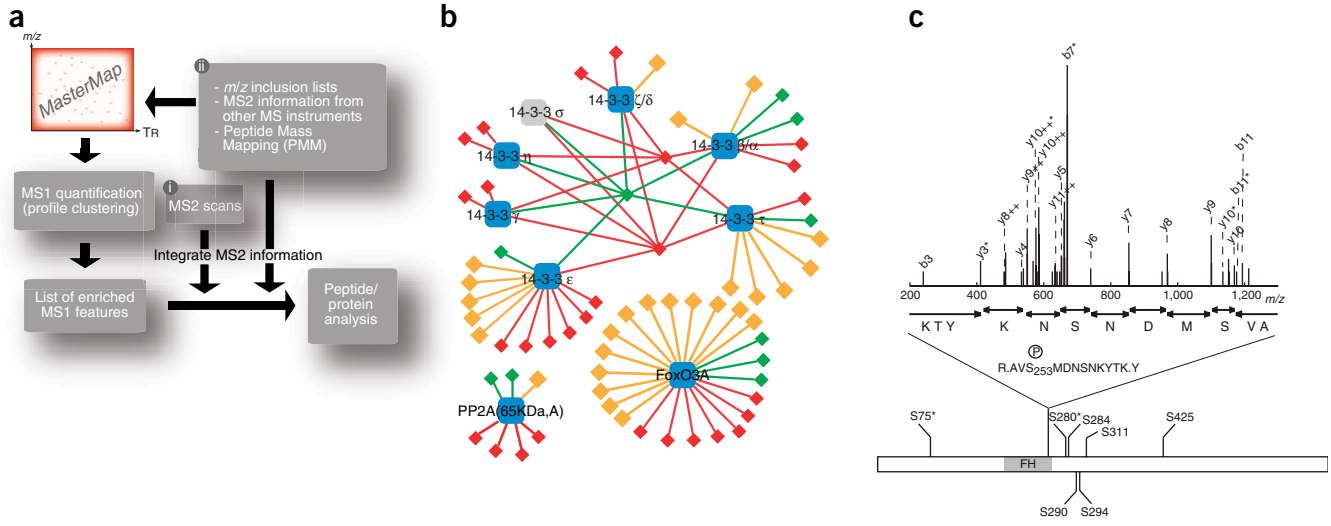

Figure 4 Extended annotation of MS1 features and peptide-to-protein associations. (**a**) MS1 features in the MasterMap are subject to quantification (here, *k*-means profile clustering). Accordingly, a list of enriched features are associated to MS2 scans acquired during the LC-MS runs (i) or further annotated by *m*/*z* inclusions lists, MS2 data from other MS instruments and peptide mass mapping (ii). MS1 feature annotations are stored in the MasterMap and used for peptide or protein analysis. (**b**) The problem of confirming the presence of a protein by MS2 information is illustrated for a single cross-linking experiment under LY-conditions. Peptides (small diamonds) and proteins (large blue and gray rectangles) are represented as nodes in a Cytoscape[42] network structure, where edges reflect the association of peptide to a protein. Peptides that are connected to a protein with a single spoke are proteotypic peptides associated with a single protein isoform. Peptides that connect more than one protein node cannot differentiate between two or more proteins. Different levels of peptide identifications are obtained either by MS2 scans, inclusions lists (green) or by PMM (red). Although the presence of 14-3-3 δ (gray) cannot be confirmed unambiguously, the other proteins (blue) are detected by proteotypic peptides. (**c**) Phosphopeptides of FoxO3A identified by tandem mass spectrometry in random sequencing mode and targeted MS2 of candidate peptides with characteristic mass shifts. The site S75 and S280 have not been described before. Blow-up shows the fragmentation pattern (MS2 spectrum) of a S253 phosphopeptide, which can be phosphorylated by PKB. The y-ion series is shown on the reversed sequence. Stars indicate ions with neutral loss of –98 Da. FH, forkhead domain.

β/α and τ were directly identified by MS2 in the initial LC-MS runs, 14-3-3 γ and η were only detected in this specific experiment by combining identifications from inclusion list experiments, LTQ data and PMM analysis.

Overall, the extended MS1 annotation efforts by targeted sequencing, the integration of MS2 information from LTQ runs and PMM substantially improved the peptide and protein information content in the MasterMap. The principle of selecting feature candidates is generic and could be based on other criteria such as differential feature intensities in comparative studies or features pairs with specific *m/z* shifts corresponding to post-translational modifications or isotopically labeled peptides.

Signaling-dependent changes of FoxO3A phosphorylation

Similarly to the quantification of protein interaction changes between the LY and FCS conditions, we analyzed the FoxO3A phosphorylation pattern by calculating enrichment ratios for every phosphorylated peptide and its unmodified counterpart detected in the MasterMap. Phosphorylation sites were confirmed by MASCOT searches[30] of tandem mass spectra and by the occurrence of mass shifts between MS1 features characteristic for phosphorylation ($\Delta Mr = 79.966$ or 159.932 Da) between MS1 features.

Overall, eight unique FoxO3A phosphorylation sites were detected (**Fig. 4c** and **Supplementary Table 2** online). Besides the known protein kinase B (PKB) target site S253, seven additional phosphorylation sites were identified, of which five had been previously described in the context of stress-dependent FoxO3A regulation[23]. Phosphorylation of S75 and S280 was not described previously. In general, the peptide enrichment ratios are based on observations of single peptides in, at most, two charge states, which does not permit statistical analysis. Nonetheless, inspection of the dilution profile of S253 showed that this peptide was present only in the FCS condition in contrast to its unmodified counterpart (LY/FCS ratio = 0.63). S253 has been reported to be phosphorylated by PKB[26,31], which is consistent with the calculated enrichment factor that we have measured under growth-stimulating conditions.

DISCUSSION

We developed a mass spectrometric and computational method in which simple one-step Co-IPs are used to determine specific interaction partners of protein complexes under varying cellular states. This method is based on the alignment of LC-MS patterns from control and bait Co-IP's followed by clustering of artificially created peptide dilution profiles. Differentiation between bait-specific interaction partners and unspecific background proteins is indispensable for the assembly of high-quality protein interaction networks from Co-IP data. Filtering based on MS2 identifications alone has been proposed as a quantitative measure (spectral counting[32]) and MS2-based filtering has been applied in a large-scale protein interaction study where statistical analysis can be applied to a large number of MS measurements[33]. However, this approach is not reliable when dealing with a small number of samples. The main reason for this is MS2 undersampling of the MS1 features. Although isotopic labeling of bait and control proteins, for example with ICAT[10] or iTRAQ[11], resolves some of these issues[34], the strategy still depends upon successful MS/MS identification and quantification of peptides in a sample. As a MasterMap contains all extracted MS1 features that are above the sensitivity threshold of the mass spectrometer together with their intensity profiles, our approach has the potential to assess protein interactions more comprehensively. Dilution profiling circumvents the known problem associated with the definition of MS1 ion current ratios between two samples when a feature is only detected in one sample. In these cases, it is not possible to decide whether a peptide is really not present at detectable levels in a sample or whether the feature extraction algorithm failed. Using the dilution approach, absence of a peptide in one sample generates a very specific dilution profile, whereas a failure in feature picking results in an inconsistent profile that can easily be identified. The specific shape of the dilution profile is valid for a wide range of peptide intensities (see **Supplementary Fig. 6** online).

Quantified MS1 features in a MasterMap can be annotated using bioinformatics methods such as PMM or post-translational modification prediction to confirm the presence of specific peptides. Interesting candidate MS1 features can then be subjected to targeted MS2 analysis in another experiment. Some aspects of this approach are similar to the accurate mass-and-time tag method, where MS1 signals are compared to a database of identified peptides using their accurate masses and normalized retention times[35].

Our approach enabled us to clearly recapitulate the known interaction between Foxo3A and 14-3-3 protein ζ/δ[26] and to further show that the 14-3-3 protein family members (β/α, γ, η, ε, τ) are specifically associated with FoxO3A when PI3K signaling is active. Upon PI3K inhibition, FoxO3A moves into the nucleus—a process that correlates with lower levels of associated 14-3-3 ζ/δ[27]. This decrease in 14-3-3 ζ/δ association following PI3K inhibition was monitored quantitatively with our MS1 profiling approach. In addition, this strategy also revealed that not only 14-3-3 ζ/δ, but also the other identified 14-3-3 family members, dissociates from FoxO3A when FoxO3A moves to the nucleus after PI3K inhibition. This suggests that at least six of the seven known members of the human 14-3-3 family may contribute to nuclear shuttling of FoxO3A.

The same experimental data sets were simultaneously analyzed for both quantitative changes in the abundances of interacting proteins as well as changes in the post-translational modifications of the identified proteins. We identified several known and previously unknown phosphorylation sites by scanning the MasterMap for peptide pairs that showed mass shifts indicative of phosphorylation (**Fig. 4c** and **Supplementary Table 2** online). Quantitative analysis further revealed a positive correlation between increased FoxO3A S253 phosphorylation and increased 14-3-3 binding to FoxO3A under conditions when PKB is active. This is consistent with a mechanism whereby 14-3-3 association is regulated by PKB-dependent phosphorylation of S253.

Although this approach was applied to analyze specific protein complexes, even greater potential for the MasterMap concept lies in extending it to Co-IP experiments with multiple bait proteins. As this approach requires only low-effort, one-step purifications without labeling, it can be reliably performed with high-throughput. MasterMaps from many experiments, for example with baits chosen from a specific signaling pathway under varying signaling states, could be aligned and merged, thereby revealing higher-order protein-enrichment profiles. The presence and variations of the interaction partners of each bait would be represented in a high-dimensional MasterMap. Moreover, this map could serve as a repository that allows comparison of interaction studies between laboratories performed by different MS instruments. Besides the need for a common data representation format to store and exchange MasterMaps between different data preprocessing software, the retention time dimension could be coded by selected standard peptides or additives to the chromatography buffers, thereby allowing real-time monitoring of the solvent gradients[36].

In conclusion, we demonstrate that our approach improves specificity and sensitivity in the identification of protein-protein interactions compared with conventional Co-IP MS2–based analysis. Furthermore, the analysis of protein interactions can be extended to quantitatively map changes in protein interaction networks. The example involving FoxO3A demonstrates the usefulness of a simultaneous analysis of protein-interaction and post-translational modifications using the MasterMap approach to resolve regulatory relationships within protein interaction networks.

METHODS

MasterMap by multiple LC-MS alignment. Data analysis was performed by the program SuperHirn, which will be made publicly available on http://tools.proteomecenter.org/SuperHirn.php (Mueller, L.N. *et al.* unpublished data). The website contains also a detailed manual providing an introduction to the program usage. SuperHirn is written in C++ and is accessed via a command line interface. The program is structured into a set of modules, which contain the different LC-MS processing functionalities (LC-MS preprocessing, pairwise and multiple LC-MS alignment, feature intensity normalization, among others) and are sequentially executed during the analysis of a LC-MS experiment. The modular usage for this LC-MS data analysis was divided into two major parts; multiple alignment of acquired LC-MS runs into a MasterMap and clustering analysis of MS1 feature profiles.

LC-MS runs were preprocessed by a feature detection routine to extract peptide features from linear ion trap fourier transform ion cyclotrone resonance mass spectrometer (LTQ-FT-ICR) data, where MS1 signals of peptide features were separated from chemical noise or experimental artifacts by their isotopic distribution in the m/z dimension, and their LC elution profile in the retention time dimension. Subsequently, peptide assignments of MS2 spectra were combined with their corresponding MS1 features by the charge state (z), m/z of the precursor ion and the retention time (T_R) of the MS2 scan.

After preprocessing of the LC-MS runs, pairwise comparisons of LC-MS runs were performed to compute an alignment structure according to which the LC-MS runs were combined into a MasterMap by a multiple LC-MS alignment. SuperHirn sequentially aligned two LC-MS runs and created a new merged LC-MS run. In every merging process, the retention time of the LC-MS pair was normalized by a LC-MS alignment routine and common features were matched using the m/z, T_R and z dimension at a user-defined tolerance level ($\Delta T_R = 0.5$ min, $\Delta m/z = 0.01$Da). This process was repeated until all LC-MS runs were combined into the MasterMap. Even though SuperHirn is a generic tool to process a wide range of MS data from different MS instruments, the availability of FT high-mass precision facilitated peak matching and improved the quality of LC-MS merging. The mass tolerance for feature matching ($\Delta m/z = 0.01$Da) was chosen large enough to map almost all equal features and small enough to avoid too many false positives. At the end of the multiple alignment process, a MasterMap was created, which unified the whole input data and represented a universal and compact data repository for further quantitative analysis.

Before clustering analysis, abundance values of peptide features were normalized to reduce systematic artifacts originating from various sources as, for example, different LC injection concentrations, LC carryover and variability in ionization efficiency[21]. MS1 feature intensity normalization was performed segment-wise using a sliding window along the T_R dimension. For every retention time segment, peptide features within a T_R segment and common to all LC-MS runs were extracted and their mean abundance values $V_{F,Av}$ were computed. The ratio of each LC-MS run–specific peptide abundance value to its corresponding $V_{F,Av}$ was then determined and averaged over all common peptides features to yield a T_R segment–specific normalization coefficient for every LC-MS run.

Peptide profiling and *k*-means clustering. At this stage, the complete set of LC-MS data was archived in a preprocessed format in the MasterMap from which MS1 feature profiles were directly extracted. Missing data points in the profile were interpolated if a neighboring point was available on both sides of the missing value. To construct significant profiles, we limited profiling analysis to matched MS1 features, which had been detected at least in three LC-MS runs. Although the selection of the dilution scheme is generic, we tested different dilution outlines to obtain a dilution schema with four dilution steps that optimized the separation of contaminant proteins and bait-specific proteins (data not shown). Based on this pre-study, we decided to apply a rather conservative dilution of 10% and 20% bait/control mixing ratio that enabled a clear distinction between contaminants and specific interaction partners.

MS1 features were divided into naturally occurring profile groups according to their profile similarity by *k*-means clustering[37]. Cluster analysis was initiated with a predetermined number of k cluster centers (here $k = 10$), which were built from profiles of randomly selected features profiles. The iteration was repeated until all cluster profiles converged ($\Delta C = 10^{-15}$) or a maximal number of iteration was reached ($max = 10^5$).

A critical factor in the comparison of MS1 feature profiles is the definition of profile similarity. For *k*-means clustering, feature profiles were normalized according to equation (1) and profile similarity $PS_{i,j}$ was defined by the averaged Manhattan distance between a pair of features (i,j) (equation (2)).

$$\dot{y}_n = \frac{y_n}{\sum_{n=1}^{N_{LC/MS}} y_n} \quad (1)$$

$$PS_{i,j} = \frac{1}{N_{LC/MS}} \sum_{n=1}^{N_{LC/MS}} |\dot{y}_{i,n} - \dot{y}_{j,n}| \quad (2)$$

Profile normalization was required to uncouple similarity scores from intensity variations and allowed the clustering algorithm to group MS1 features according to their profile trends. Therefore, different peptide ionization properties or the initial protein concentrations did not dominate the calculation of similarity scores and rendered the clustering robust for true feature profiling quantification. The constructed cluster structure reduced the information of extracted peptide features profiles to a number of centroids, which reflected the major peptide abundance changes over the different LC-MS runs. Subsequent data analysis was then targeted on one or several feature subgroup(s) of interest characterized by their corresponding cluster profile(s).

Targeted protein profiling and confidence assessment. According to the experimental outline, targeted protein analysis was performed on MS1 feature members of the selected profile groups, which showed the highest similarity to the theoretical enrichment trend (target profile). Subsequently, SuperHirn reorganized MS1 features into peptides and proteins. First, charge state deconvolution was performed by grouping MS1 features according to their neutral molecular mass into decharged peptides, where additionally MS1 features containing post-translational modifications (here phosphorylation) were detected by defined molecular mass differences to their nonmodified counterpart. Integrating available information from MS2 analysis, the subset of constructed peptides containing MS1 features with high confidence MS2 identifications[24] (PeptideProphet peptide probability > 0.9) were reorganized into proteins. Protein consensus profiles were then computed by averaging over the features members of all peptides of a protein. Protein profile similarity scores $PS_{ProteinX}$ to the theoretical enrichment trend were determined as previously described (equation (2)).

Profile similarity assessment by computed $PS_{ProteinX}$ enabled evaluation of which protein consensus profiles followed closely the theoretical enrichment trend and represented candidates for interaction partners of the bait protein. However, as described for other proteomics data[38], the challenge remains to evaluate high-scoring interaction candidates and to determine whether their score reflects a true similarity to the target profile. The ProfileProphet module of SuperHirn performed a mixture analysis of obtained similarity scores and transformed protein profile scores into probability values of true similarity $p(+|PS_{ProteinX})$ to the theoretical enrichment trend. The distribution of correct and false scores was modeled by two Gaussian distributions and their mean μ and s.d. σ were determined by an expectation maximization algorithm[39]. The probability $p(+|PS_{ProteinX})$ of a true profile similarity given a score $PS_{ProteinX}$ was defined in equation (3) using Bayes' law, where $p(PS_{ProteinX}|+)$ and $p(PS_{ProteinX}|-)$ were computed from the modeled Gaussian distributions, and $p(+)$ and $p(-)$, respectively, were calculated from the mixture probability of true and false profile similarities. ProfileProphet probabilities enabled

assessment of the significance of profile similarity scores and statistical evaluation of target profile similarities of potential bait interactors.

$$p(+|PS_{ProteinX}) = \frac{p(PS_{ProteinX}|+)p(+)}{p(PS_{ProteinX}|+)p(+) + p(PS_{ProteinX}|-)p(-)} \quad (3)$$

Peptide mass mapping. An in-house software using the InSilicoSpectro[40] perl library was used to compare peptides with theoretical masses of *in silico* protein digests ($\Delta m/z = 7$ p.p.m., tryptic digestion allowing one missed cleavage, methionine oxidation was considered only if both oxidized und unmodified peptide masses were detected). Predicted retention times of theoretical protein digest (using the Petritis algorithm in InSilicoSpectro[40]) had to match the experimental retention time of the measured peptides (removing about one-third of the peptides). For the cross-linked samples, T_R could not be predicted due to cross-linker modification of the peptides and no retention time filtering was applied. Even with the high-accuracy mass measurements used, mass matching is prone to false positives owing to the large number of features detected and the number of matched features can serve only as an indication, especially if this number is <4.

Cell lines and transfections. HEK 293 cells were grown in DMEM medium (4.5 g/l glucose, 50 μg penicillin/ml, 50 μg/ml streptomycin, 10% FCS). To establish control cells and cells stably expressing human hemagglutin tagged FoxO3A, we transfected HEK 293 cells with either pcDNA3 or pcDNA3-HA-FoxO3A, using the Fugene transfection reagent (Roche). We added G418 to the medium (1 mg/ml) 48 h after transfection and stably transfected cell pools were obtained after 5 weeks of G418 selection. Expression of HA-FoxO3A in G418-resistant pools was confirmed by western blot analysis, using the monoclonal anti-HA antibody HA-11 (Covance).

Immunofluorescence microscopy. Cell pools stably expressing HA-FoxO3A seeded on glass cover slips were either grown in DMEM containing 10% FCS or were serum-starved overnight and treated with 20 μM LY294002 inhibitor (Invitrogen) for 1 h. Cells were washed once in PBS before fixation for 15 min in 3.7% paraformaldehyde/PBS. Fixed cells were washed 3× with PBS and treated for 5 min with 0.5% Triton/PBS. Following three washes with PBS, cells were incubated with a 1:10 dilution of a monoclonal antibody 12CA5 tissue culture supernatant for 1 h at 23 °C. Cells were washed 3× with PBS and stained with a 1:200 dilution of an anti-mouse-Alexa488 secondary antibody (Molecular Probes) and DAPI (4′,6-diamidino-2-phenylindole). Finally, cells were washed 3× with PBS, mounted in Vectashield (Vector Laboratories) and images of confocal sections were obtained on a Leica SP2 AOBS confocal microscope.

Immunoprecipitation. HA-FoxO3A-expressing HEK293 cells and control HEK293 cells were grown in 15-cm dishes in DMEM medium supplemented with 10% FCS in the presence of antibiotics (50 μg penicillin/ml, 50 μg/ml streptomycin; Gibco). Each cell line was split into two groups. One group was grown as before in 10% FCS (FoxO3A-FCS and control-FCS). The other group (FoxO3A-LY and Control-LY) was serum-starved overnight and treated for 1 h with LY294002 inhibitor (20 μM final concentration) before harvesting the cells. Ten dishes per group ($\sim 3 \times 10^8$ cells) were harvested, rinsed once with PBS and lysed in 10 ml of ice-cold TNN-HS (Tris NaCl Nonidet P-40 High Salt) buffer (0.5% NP-40, 50 mM Tris, pH 7.5, 250 mM NaCl, 1 mM EDTA, 50 mM NaF, 1.5 mM Na_3VO_4, 1 mM DTT, 0.1 mM PMSF, 1 protease inhibitor tablet mix (Roche) per 50 ml). Cells were lysed 10 min. on ice with ten strokes using a tight-fitting Dounce homogenizer.

Insoluble material was removed by centrifugation. The cleared extracts were precleared with 50 μl ProteinA-Sepharose (Amersham) for 1 h. After removal of ProteinA-Sepharose 50 μl of anti HA-(12CA5) monoclonal antibodies covalently coupled to ProteinA-Sepharose was added to each of the extracts and incubated for 4 h with gentle rocking at 4 °C. Immunoprecipitates were washed 3× with ten bead volumes of the lysis buffer on spin columns (Biorad) and twice with ten-bead volumes of the corresponding lysis buffer without protease inhibitor and without detergent.

Proteins were eluted from the beads with three bead volumes of 0.2 M glycine buffer (pH 2.3) and neutralized subsequently with 100 mM NH_4HCO_3. Cystine bonds and DSP-linked peptides were reduced in 5 mM TCEP for 30 min. at 37 °C and alkylated in 10 mM iodocetamide for 30 min. at 23 °C , respectively. The samples were digested overnight with 1 μg trypsin (Promega) each. Peptides were purified with ultramicro spin columns (Harvard Apparatus), according to the manufacturers protocol, and redissolved in 0.1% formic acid for injection into the mass spectrometer.

LC-MS analysis. LC-MS analysis of Co-IP samples was performed on a Thermo Fourier Transformed-LTQ mass spectrometer (Thermo Electron), which was connected to an electrospray ionizer. TheAgilent chromatographic separation system 1100 (Agilent Technologies) was used for peptide separation. The LC system was connected to a 10.5-cm fused silica emitter of 150 μm inner diameter (BGB Analytik) and was packed by in-house Magic C18 AQ 5 μm resin (Michrom BioResources). The Agilent auto sampler was used to load samples at a temperature of 6 °C. Subsequently, the sample was separated during 70 min with a linear gradient of a 5–40% acetonitrile/water mixture, containing 0.1% formic acid, at a flow rate of 1.2 μl/min. After every LC-MS analysis, the LC system was washed with 50% trifluoroethanol to prevent cross-contamination between the different analyzed samples and a LC-MS run of a 200 fmol GluFib standard protein sample was acquired to monitor the chromatographic performance.

The data acquisition mode was set to obtain FT-MS1 scans at a resolution of 100,000 full width at half maximum (at m/z 400), where each MS1 scan was followed by three MS2 scans in the linear ion trap (overall cycle time of ~ 1 s.). To increase the efficiency of MS2 attempts, the charged state screening modus of the FT-MS was used to exclude unassigned or singly charges ions. For every MS1 scan, data dependent aquisition was performed on the three most intense ions if the ion count exceeded a threshold of 200 ion counts.

MS2 peptide assignments. Acquired MS2 scans were searched against the human International Protein Index (IPI) protein database (v.3.15) using the SORCERER-SEQUEST (TM) v3.0.3 search algorithm, which was run on the SageN Sorcerer (Thermo Electron). *In silico* trypsin digestion was performed after lysine and arginine (unless followed by proline) tolerating two missed cleavages in fully tryptic peptides. Database search parameters were set to allow phosphorylation (+79.9663 Da) of serine, threonine and tyrosine as a variable modification and carboxyamidomethylation (+57.021464 Da) of cysteine residues as fixed modification. For the searches of DSP cross-linked samples an additional variable modification of lysine residues (+145.01975) from the carboxyamidomethylated cleaved DSP cross-linker was considered. Search results were evaluated on the Trans Proteomic Pipeline (TPP)[41] using Peptide Prophet (v3.0)[24].

To specifically identify phospho-peptides, we searched MS2 spectra with the MASCOT search engine[30] against the human IPI database version 1.5, which considers neutral phosphate loss fragments. Searches were performed on fully tryptic peptides allowing for two missed cleavages. Carboxiamidomethylated cysteines were set as a fixed modification and phosphorylation of serine, threonine and tyrosine were set as variable modifications. The fragment mass tolerance was set to 0.5 Da, precursor mass tolerance to 10 p.p.m. The instrument type was set as FT-ICR.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors thank Luis Mendoza for helpful discussions. Alexander Schmidt for help with the FT-ICR spectrometer. We thank M. Greenberg for the gift of pcDNA3-HA-FoxO3A. The work was funded in part by ETH Zurich with Federal (USA) funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract No. N01-HV-28179. Oliver Rinner was supported by fellowships of the Roche Research Foundation and the Deutsche Forschungsgemeinschaft (DFG).

AUTHOR CONTRIBUTIONS

O.R. conducted most of the experimental work and developed analytical concepts, L.N.M. conceptualized and implemented most of the SuperHirn software, M.H. did most of the LTQ-FT-ICR measurements, M.M., M.G. and R.A. shared senior authorship responsibilities and have conceived computational, biological and mass spectrometric concepts, respectively.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Ito, T., Tashiro, K. & Kuhara, T. Systematic analysis of *Saccharomyces cerevisiae* genome: gene network and protein-protein interaction network. *Tanpakushitsu Kakusan Koso* **46**, 2407–2413 (2001).
2. Uetz, P. & Hughes, R.E. Systematic and large-scale two-hybrid screens. *Curr. Opin. Microbiol.* **3**, 303–308 (2000).
3. Gavin, A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
4. Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
5. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
6. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
7. Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
8. Figeys, D. Functional proteomics: mapping protein-protein interactions and pathways. *Curr. Opin. Mol. Ther.* **4**, 210–215 (2002).
9. Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
10. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
11. Ross, P.L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
12. Schmidt, A., Kellermann, J. & Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4–15 (2005).
13. Ong, S.E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
14. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **96**, 6591–6596 (1999).
15. Ranish, J.A. *et al.* The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* **33**, 349–355 (2003).
16. Himeda, C.L. *et al.* Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. *Mol. Cell. Biol.* **24**, 2132–2143 (2004).
17. Blagoev, B. *et al.* A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* **21**, 315–318 (2003).
18. Brand, M. *et al.* Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat. Struct. Mol. Biol.* **11**, 73–80 (2004).
19. Li, X.J., Yi, E.C., Kemp, C.J., Zhang, H. & Aebersold, R. A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics* **4**, 1328–1340 (2005).
20. Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P. & Geromanos, S.J. Absolute quantification of proteins by LCMS: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5**, 144–156 (2006).
21. Callister, S.J. *et al.* Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **5**, 277–286 (2006).
22. Andersen, J.S. *et al.* Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574 (2003).
23. Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).
24. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
25. Van Der Heide, L.P., Hoekman, M.F. & Smidt, M.P. The ins and outs of FoxO shuttling: mechanisms of FoxO translocation and transcriptional regulation. *Biochem. J.* **380**, 297–309 (2004).
26. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
27. Brunet, A. *et al.* 14–3-3 transits to the nucleus and participates in dynamic nucleo-cytoplasmic transport. *J. Cell Biol.* **156**, 817–828 (2002).
28. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
29. Nesvizhskii, A.I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **4**, 1419–1440 (2005).
30. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
31. Kops, G.J. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
32. Liu, H., Sadygov, R.G. & Yates, J.R., III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
33. Gavin, I.M., Kukhtin, A., Glesne, D., Schabacker, D. & Chandler, D.P. Analysis of protein interaction and function with a 3-dimensional MALDI-MS protein array. *Biotechniques* **39**, 99–107 (2005).
34. Guerrero, C., Tagwerker, C., Kaiser, P. & Huang, L. An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26S proteasome-interacting network. *Mol. Cell. Proteomics* **5**, 366–378 (2006).
35. Smith, R.D. *et al.* An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2**, 513–523 (2002).
36. Chen, S.S. & Aebersold, R. LC-MS solvent composition monitoring and chromatography alignment using mobile phase tracer molecules. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **829**, 107–114 (2005).
37. Duda, R.O., Hart, P.E. & Stork, D.G. *Pattern Classification* vol. 1, edn. 2, (Wiley-Interscience, New York, 2000).
38. Nesvizhskii, A.I. & Aebersold, R. Analysis, statistical validation and dissemination of large-scale proteomics data sets generated by tandem MS. *Drug Discov. Today* **9**, 173–181 (2004).
39. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, New York, 2001).
40. Colinge, J., Masselot, A., Carbonell, P. & Appel, R.D. InSilicoSpectro: an open-source proteomics library. *J. Proteome Res.* **5**, 619–624 (2006).
41. Keller, A., Eng, J., Zhang, N., Li, X.J. & Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **1** 2005, 0017 (2005).
42. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).