



# Reconstruction and analysis of nutrient-induced phosphorylation networks in *Arabidopsis thaliana*

Guangyou Duan<sup>1</sup>, Dirk Walther<sup>1\*</sup> and Waltraud X. Schulze<sup>1,2\*</sup>

<sup>1</sup> Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

<sup>2</sup> Department of Plant Systems Biology, Universität Hohenheim, Stuttgart, Germany

## Edited by:

Sonia Osorio, Málaga University, Spain

## Reviewed by:

Wolfram Weckwerth, University of Vienna, Austria

Jedrezej J. Szymanski, Max Planck Institute of Molecular Plant Physiology, Germany

## \*Correspondence:

Dirk Walther, Max Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Brandenburg, Germany  
e-mail: walther@mpimp-golm.mpg.de

Waltraud X. Schulze, Fg. Systembiologie der Pflanze, Universität Hohenheim, Garbenstr. 30, BIO I, Laborbau, 292, 70599 Stuttgart, Germany  
e-mail: wschulze@uni-hohenheim.de

Elucidating the dynamics of molecular processes in living organisms in response to external perturbations is a central goal in modern systems biology. We investigated the dynamics of protein phosphorylation events in *Arabidopsis thaliana* exposed to changing nutrient conditions. Phosphopeptide expression levels were detected at five consecutive time points over a time interval of 30 min after nutrient resupply following prior starvation. The three tested inorganic, ionic nutrients  $\text{NH}_4^+$ ,  $\text{NO}_3^-$ ,  $\text{PO}_4^{3-}$  elicited similar phosphosignaling responses that were distinguishable from those invoked by the sugars mannitol, sucrose. When embedded in the protein–protein interaction network of *Arabidopsis thaliana*, phosphoproteins were found to exhibit a higher degree compared to average proteins. Based on the time-series data, we reconstructed a network of regulatory interactions mediated by phosphorylation. The performance of different network inference methods was evaluated by the observed likelihood of physical interactions within and across different subcellular compartments and based on gene ontology semantic similarity. The dynamic phosphorylation network was then reconstructed using a Pearson correlation method with added directionality based on partial variance differences. The topology of the inferred integrated network corresponds to an information dissemination architecture, in which the phosphorylation signal is passed on to an increasing number of phosphoproteins stratified into an initiation, processing, and effector layer. Specific phosphorylation peptide motifs associated with the distinct layers were identified indicating the action of layer-specific kinases. Despite the limited temporal resolution, combined with information on subcellular location, the available time-series data proved useful for reconstructing the dynamics of the molecular signaling cascade in response to nutrient stress conditions in the plant *Arabidopsis thaliana*.

**Keywords:** phosphorylation, protein–protein interaction, signaling transduction, network inference, correlation, *Arabidopsis thaliana*

## INTRODUCTION

Intra-cellular communication and information processing is performed by complex, dynamic, and context-specific molecular signaling networks (Terfve and Saez-Rodriguez, 2012). Protein phosphorylation is the best known post-translational modification involved in molecular signaling and plays a significant role in a wide range of cellular processes in all organisms (Cohen, 2000; Macek et al., 2008; Schulze, 2010). Enabled by the development of novel sensitive experimental techniques to detect phosphorylation sites in proteins (Larsen et al., 2005; Sugiyama et al., 2007), the principle of phosphorylation-mediated activation and inactivation of proteins and the modulation of molecular interactions has been studied intensively in a broad range of organisms (Pawson, 1997; Pawson and Nash, 2003; Pawson and Taylor, 2009). Time course studies of phosphorylation events are of particularly high value in the context of signaling as they allow to reveal the dynamics and the cause–effect relationships between all molecular components involved in the signal transduction process (Blagoev et al., 2004; Olsen et al., 2006; Niittylä et al., 2007; Engelsberger and Schulze, 2012).

Besides advances on the instrumentation side, computational methods have been devised to interrogate the resulting data sets. Both descriptive as well as predictive computational methods have been developed [For review, see Janes and Yaffe (2006)]. In the category of *descriptive* approaches, in particular the methods clustering (Olsen et al., 2006; Huang et al., 2007; Engelsberger and Schulze, 2012), principal component analysis (PCA), singular value decomposition (SVD), data mapping onto protein interaction networks (PIN), and protein signaling networks (PSN) (Krüger et al., 2008; Jørgensen et al., 2009) have been applied with the aim to reduce the complex behavior of a molecular system to the core components, thereby “describing” it in the most meaningful and parsimonious way. The primary goal of descriptive approaches is to identify common as well as distinguishing properties of the system under study exposed to different conditions or at different time points. By contrast, *predictive* approaches allow predicting the system’s response given a set of conditions and data for appropriate independent predictor variables (Terfve and Saez-Rodriguez, 2012). Predictive approaches employ input/output regression based methods (such as partial

least squares regression (Gaudet et al., 2005; Janes et al., 2005), multiple linear regression (Ekins et al., 2008; Alexopoulos et al., 2010) as well methods that aim to infer (in a sense “predict”) the web of interactions between all components based on the available data using network inference methods [such as correlation based inference methods (Ciaccio et al., 2010), modular response analysis (Kholodenko, 2006; Santos et al., 2007), multiple input multiple output models (Nelander et al., 2008), Bayesian network inference (Sachs et al., 2005; Ciaccio et al., 2010)] or reaction-based models such as ordinary and partial differential equations (Aldridge et al., 2006; Birtwistle and Kholodenko, 2009; Chen et al., 2009), rule-based models (Borisov et al., 2005; Conzelmann et al., 2006; Hlavacek et al., 2006; Danos et al., 2007; Faeder et al., 2009; Feret et al., 2009; Sneddon et al., 2011), and logic-based models (Gat-Viks and Shamir, 2007; Watterson et al., 2008; Saez-Rodriguez et al., 2009; Morris et al., 2010, 2011; Schlatter et al., 2011).

The field of network reconstruction gained particularly strong traction in the context of gene expression regulation, where reverse engineering models have been designed to infer gene regulatory networks from gene expression data (Bansal et al., 2007; Hecker et al., 2009). More recently, efforts were undertaken to use the generated large-scale experimental phosphorylation site data that are now available in public databases to reconstruct kinase-specific phosphorylation interactions, many of which also make use of “known” protein–protein interactions (Linding et al., 2007; Xue et al., 2008; Song et al., 2012; Newman et al., 2013).

In the plant research field, phosphoproteomic experiments interrogating the dynamics of phosphorylation events by capturing more than two time points are still very scarce (Niittylä et al., 2007; Chen et al., 2010; Engelsberger and Schulze, 2012). By contrast, pairwise comparisons of two conditions have been carried out rather frequently (Benschop et al., 2007; Li et al., 2009; Reiland et al., 2009; Kline et al., 2010).

As sessile organisms, plants have evolved sensitive mechanisms in their plasma membrane to detect and respond to rapid changes in external nutrient conditions [e.g., reviewed for responses to nitrate in Wang et al. (2012b)]. As still very little is known about post-translational regulation of nutrient-induced signaling processes, these phosphoproteomic experiments complement the existing knowledge gained from the analysis of nutrient-induced transcript changes (Scheible et al., 1997; Morcuende et al., 2007; Osuna et al., 2007; Krouk et al., 2010).

In this study, phosphoproteomics data obtained from starvation-resupply experiments involving several different nutrient conditions (nitrate, phosphate, ammonium, and the sugars mannitol and sucrose) and sampling at five consecutive time points (Niittylä et al., 2007; Engelsberger and Schulze, 2012) were analyzed. In the published experiments, mannitol has served as osmotic control and was also included here to define osmotic responses associated with nutrient changes. Based on the time-resolved phosphoproteomic data set, we conducted a systematic computational analysis of the dynamics of the observed in nutrient-induced phosphorylation events by applying descriptive approaches including clustering, data mapping onto PINs as well as predictive approaches resulting in dynamic phosphorylation

network reconstructions. The objectives of this study were (a) to identify commonalities as well as characteristic differences of the phosphosignaling in response to different nutrient responses, and (b), to exploit the available time-series dataset to test various network reconstruction methods and performance metrics for their suitability to generate plausible networks even if only short time series data serve as input.

Our results demonstrate that current proteomics technologies allow monitoring of dynamic phosphorylation cascades at sufficient resolution and precision amenable for an application of computational network inference methods. The identified networks are characterized by distinct topological layers involved in signal initiation, processing, and an effector level. With time progression, we observed an increasingly broadened recruitment of phosphoproteins characteristic of an information dissemination flow in response to nutrient stimulation. Specific phosphorylation peptide motifs associated with them were identified indicating the action of layer-specific kinases.

## RESULTS

The dataset used in this study consists of phosphopeptide identifications from starvation-resupply time course experiments (Niittylä et al., 2007; Engelsberger and Schulze, 2012) and is available from the PhosPhAt database (Durek et al., 2010). Label-free ion intensity quantitation including retention time alignment was carried out to obtain quantitative dynamic information across five time points of phosphopeptide abundance upon resupply to starved seedlings with five nutrients/organic substances ( $\text{NH}_4^+$ ,  $\text{NO}_3^-$ ,  $\text{PO}_4^{3-}$ , sucrose, and mannitol). As most network reconstruction methods rely on the data to be complete, only datasets with quantitative information for all five time points were retained for further analysis (Table 1). In total, 546 different peptide sequences were used in the analysis.

### CHARACTERISTICS OF THE CONDITION-SPECIFIC PHOSPHOPROTEIN DATA SETS

#### *Phosphoprotein-set comparison across different nutrient conditions*

We first aimed to characterize the phosphorylation events following different nutrient and osmotic challenges based on the presence and absence of phosphorylated proteins. Thus, initially we ignored the dynamics of any changes of the phosphorylation state of individual phosphorylation sites. Some nutrient/osmotic treatments exhibited a greater than expected overlap of their respective phosphoprotein set indicative of similar signaling response processes (Table 2, File S5). For example, while only one protein would be expected to be observed in common between  $\text{NO}_3^-$  and  $\text{NH}_4^+$  treatment when drawn randomly, 19 phosphoproteins were

**Table 1 | Number of phosphorylated peptides and proteins in each nutrient/organic solute starvation-resupply treatment with complete quantitative information across all five time points.**

	Mannitol	$\text{NH}_4$	$\text{NO}_3$	$\text{PO}_4$	Sucrose
Peptides	180	76	86	144	60
Proteins	157	70	75	129	53

**Table 2 | Number of phosphoproteins found in common between two treatments.**

	Mannitol	NH <sub>4</sub>	NO <sub>3</sub>	PO <sub>4</sub>	sucrose
Mannitol	157 (4.71)	3 (2.10)	5 (2.25)	13 (3.87)	13 (1.59)
NH <sub>4</sub>	3 (2.10)	70 (0.94)	19 (1.00)	13 (1.72)	5 (0.71)
NO <sub>3</sub>	5 (2.25)	19 (1.00)	75 (1.07)	22 (1.85)	8 (0.76)
PO <sub>4</sub>	13 (3.87)	13 (1.72)	22 (1.85)	129 (3.18)	9 (1.31)
sucrose	13 (1.59)	5 (0.71)	8 (0.76)	9 (1.31)	53 (0.54)

The number in parentheses is the expected common number of proteins in each pair of treatments based on the PhosPhAt database.

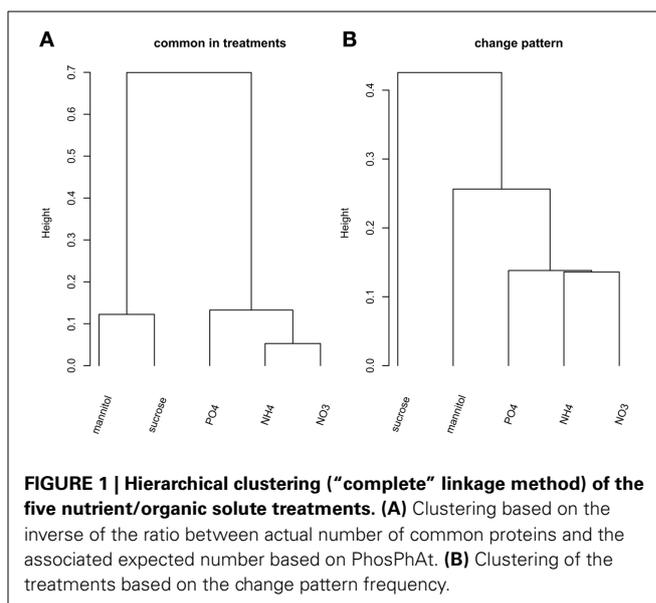
actually detected under both conditions. By contrast, other treatments are characterized by less similar phosphoprotein sets pointing to distinct phosphoprotein-signaling events, such as under the osmotic mannitol treatment and NH<sub>4</sub><sup>+</sup> nutrient treatment conditions. Here, the number of commonly observed phosphoproteins is close to random. The overlap of the commonly identified proteins for all nutrient conditions is visualized by hierarchically clustering the inverse of the ratio between actually observed common phosphoproteins in both of the considered treatment conditions relative to the corresponding expected random number as the distance metric. Smaller values of the inverse indicate larger overlaps, and thus, similar phosphoprotein sets. The different treatments segregated into two main classes, one containing treatments with the organic molecules mannitol and sucrose, the other containing the treatment with inorganic ions/salts NH<sub>4</sub><sup>+</sup>, NO<sub>3</sub><sup>-</sup>, and PO<sub>4</sub><sup>3-</sup> (Figure 1A).

#### Dynamics of stimulus induced phosphorylation

Next, we included the dynamic information into the comparison of substance-induced phosphorylation events by focusing first on the qualitative change patterns across time. The *Change pattern* describes qualitatively how phosphorylation levels change from one time point to the next for a given phosphopeptide [increase, decrease, or unchanged relative to defined thresholds (see Materials and Methods), File S1]. The clustering of the different treatment conditions based on the change pattern statistic associated with all detected, nutrient-condition-specific phosphopeptides (Figure 1B) resembles the clustering obtained from the static presence/absence call (Figure 1A). Again, the inorganic nutrients (NO<sub>3</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>, PO<sub>4</sub><sup>3-</sup>) cluster together and are distinct from the organic solute treatments (mannitol and sucrose) with the latter exhibiting differing change patterns.

#### PHOSPHORYLATION DYNAMICS FROM A NETWORK PERSPECTIVE

So far, we have analyzed the available dynamic phosphorylation data in a statistically-descriptive and comparative fashion, in which the set of phosphoproteins were treated as independent entities. However, in the real biological system they are part of an interaction network and act in a whole-cell context. Thus, we investigated the set of phosphoprotein when embedded into the context of known protein-protein interactions and, secondly, we aimed to reconstruct the dynamic regulatory relationships between the identified phosphoproteins directly from the data by applying network inference methods.



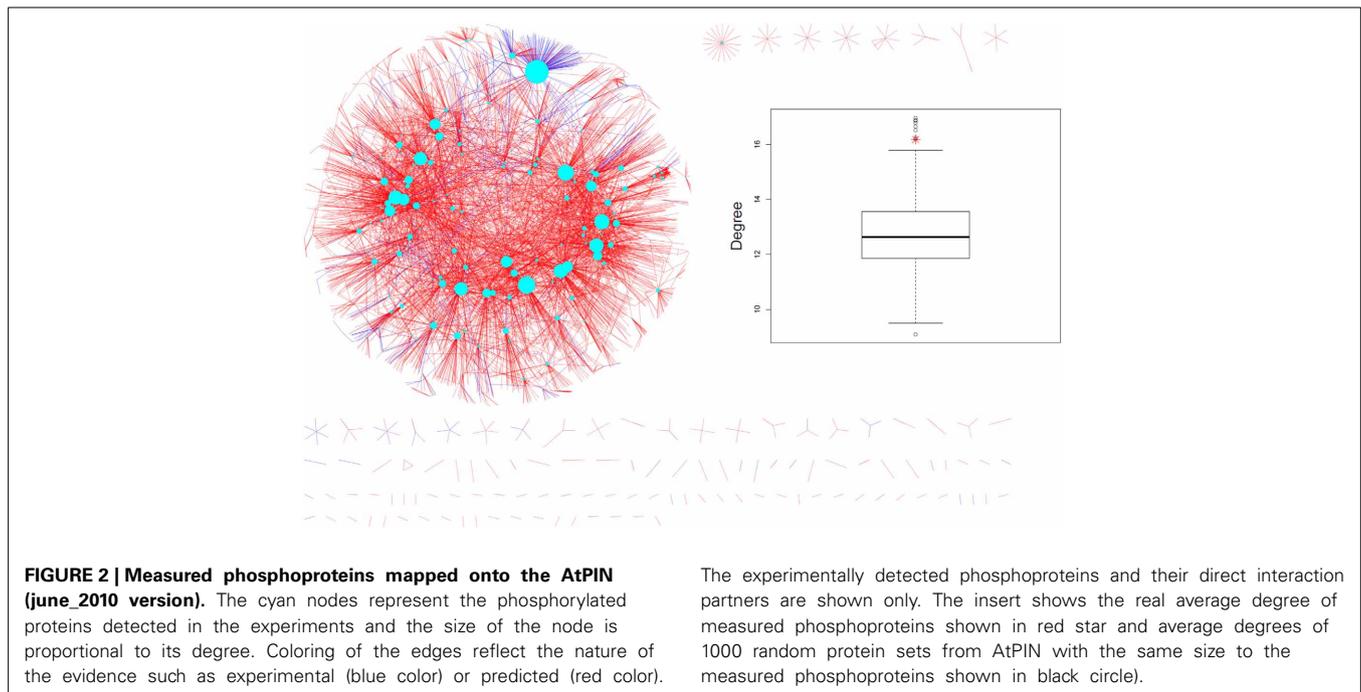
**FIGURE 1 | Hierarchical clustering (“complete” linkage method) of the five nutrient/organic solute treatments. (A)** Clustering based on the inverse of the ratio between actual number of common proteins and the associated expected number based on PhosPhAt. **(B)** Clustering of the treatments based on the change pattern frequency.

#### Mapping phosphorylated proteins onto protein interaction networks

Since phosphorylation events are associated with protein-protein interactions via physical kinase-substrate encounters, we mapped the set of experimentally observed phosphoproteins onto the PIN of *Arabidopsis thaliana* (Dreze et al., 2011). This provided (a) context for those proteins that were observed in our dataset but without detecting their interaction partners, and (b) allowed us to investigate the specific role of phosphoproteins within the global network of physical interactions. We used AtPIN (Brandão et al., 2009) as a large-scale reference PIN, which, in addition to experimentally identified interactions, also includes many predicted interactions based on annotation transfer.

When mapped onto the AtPIN (Figure 2), the experimentally identified phosphoproteins were found to possess significantly higher interaction degrees [median (mean) of 6 (16.18)] compared to all other proteins [median (mean) of 4 (12.69),  $p$ -value = 1E-04 according to the non-parametric Mann-Whitney rank sum test (MWT)]. As our dataset represents only a small fraction of all experimentally identified phosphoproteins in Arabidopsis, we extended the PIN-degree comparison to all phosphoproteins contained in PhosPhAt (Durek et al., 2010) and compared them to all other proteins in the AtPIN. Again, phosphoproteins were found to engage in more physical interactions [median (mean) of 5 (15.28)] than non-phosphorylated proteins median (mean) of 3 (12.05),  $p$ -value<sub>MWT</sub> = 0]. The average degree of phosphorylated proteins was found significantly higher also compared to a series of 1000 same-sized random protein sets drawn from the AtPIN (Insert in Figure 2).

We checked whether protein abundance may be a confounding factor in our analysis. While significant, no relevant positive correlation was observed between degree and reported protein abundance (Piques et al., 2009) with a Spearman rank correlation of  $r = 0.086$  ( $p = 0.0001$ ). Thus, protein abundance associated with an increased likelihood of experimental detection and



possible increased degree cannot explain the observed degree differences. Thus, phosphoproteins appear to play a central role in the network of physical interactions between the proteins in Arabidopsis.

### Network inference using a novel scoring scheme

When reconstructing networks from OMICS data, it is challenging to evaluate the accuracy of the employed reconstruction method without availability of suitable benchmark networks. Since the dataset studied here comprises only five time points, network inference is particularly challenging as the statistical power is low. We applied two scoring schemes (see Materials and Methods) that employ prior knowledge to help identify the most suitable inference method. The first method relied on subcellular localization information and scored predicted interactions by how likely they are based on the general observed frequency of interactions between proteins from the same or different compartments (SLPF-score, see Materials and Methods). The second scheme used the semantic similarity method (Wang et al., 2007) and judged predicted interactions based on the similarity of the respective Gene Ontology annotation terms as interacting proteins are likely involved in the same biological process and function.

A range of different reconstruction methods were tested and ranked based on the two scoring schemes. The applied inference methods included relevance based methods (Pearson correlation, first order partial correlation, full order partial correlation), mutual information based methods (MI), graphical model methods (Dynamic Bayesian network, DBN). Based on the two scoring schemes, Pearson correlation performed better than partial correlation (Table 3). The relevance based methods (correlation-based methods) also performed better than model- and mutual information-based methods, probably explained by

the limitations of the dataset (short length and noise). It should be noted that the GO annotation used in the semantic similarity score was possibly based in part on protein interaction information as available in AtPIN as well so that the two applied scoring schemes are not entirely independent.

Based on the overall-performance of the various inference methods as judged by the two scoring schemes (Table 3), we decided to use the Pearson correlation-based method to produce the final network of phosphoproteins derived from the available data. As a disadvantage compared to GGMs or other graphical model methods, Pearson correlation-based inference methods result in undirected networks. In order to infer cause-effect relationships (i.e., directed networks), we assigned directionality to every edge as introduced in (Schäfer and Strimmer, 2005; Opgen-Rhein and Strimmer, 2007). Here, directionality assignments are based on the log ratio of partial variance between the two nodes connected by an edge (see Materials and Methods).

The different nutrient/osmotic treatments resulted in separate subnetworks connected, and thus merged, via central proteins involved in compound uptake and signaling (Figure 3). For example, plasma membrane ATPases (AT2G18960, AT4G30190), and water channel PIP2E (AT2G39010) are regulated by phosphorylation and are involved in all different treatments. Among the signaling proteins, calmodulin binding protein (AT1G74690) was detected involved in the organic molecule responses (mannitol and sucrose), but not with the inorganic salts. Three different nutrient responses involved the MAP kinase MKK2 (AT4G29810). In all treatments, phosphorylation changes were observed for the metabolic enzyme glutamate synthase (AT1G66200), which was reported to integrate responses for changes in carbon, nitrogen, and phosphate status through alterations in phosphorylation status (Oliveira and Coruzzi, 1999; Lima et al., 2006).

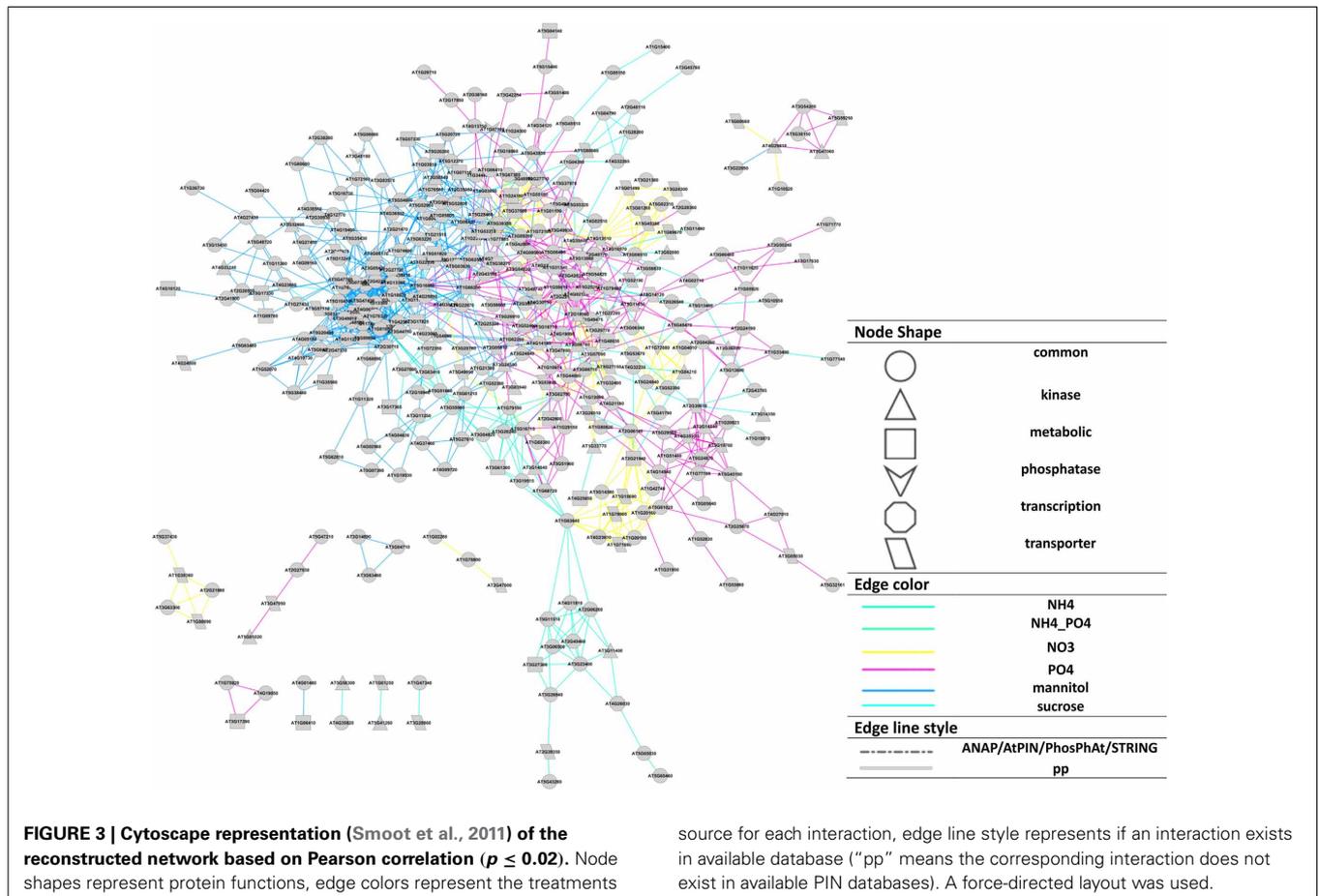
**Table 3 | Accuracy scores of the reconstructed networks with same size generated by different methods using the subcellular location pair frequency (SLPF) metric and semantic similarity of gene ontology terms (BP, CC, and MF).**

	DBN	FirstOrderPartialCorrelation	FullOrderPartialCorrelation	MI	PearsonCorrelation
BP_Jiang	0.4958	0.4996	0.4986	0.4930	0.5054
BP_Lin	0.2730	0.2940	0.2895	0.2915	0.2994
BP_Rel	0.1993	0.2142	0.2146	0.2132	0.2132
BP_Resnik	0.0909	0.0968	0.0987	0.0984	0.0978
BP_Wang	0.4416	0.4311	0.4391	0.4424	0.4340
CC_Jiang	0.7569	0.7763	0.7704	0.7775	0.7791
CC_Lin	0.5148	0.5463	0.5407	0.5561	0.5619
CC_Rel	0.4142	0.4482	0.4386	0.4584	0.4667
CC_Resnik	0.1235	0.1316	0.1294	0.1344	0.1373
CC_Wang	0.6303	0.6396	0.6390	0.6423	0.6434
MF_Jiang	0.3178	0.3388	0.3399	0.3213	0.3170
MF_Lin	0.1957	0.2067	0.2046	0.2118	0.2051
MF_Rel	0.1625	0.1747	0.1705	0.1837	0.1741
MF_Resnik	0.0833	0.0904	0.0881	0.0988	0.0942
MF_Wang	0.4218	0.4188	0.4339	0.4075	0.4085
SLPF score	0.1299	0.1221	0.1195	0.1360	0.1368

“\*\_Jiang”, “\*\_Lin”, “\*\_Rel”, “\*\_Resnik” and “\*\_Wang” represent different semantic similarity metric as introduced in Wang et al. (2007).

Larger score values correspond to networks with higher confidence. The gray shaded values identify the highest score within each scoring metric.

DBN, dynamic Bayesian network; FirstOrderPartialCorrelation, first order partial correlation; FullOrderPartialCorrelation, full order partial correlation; MI, mutual information; PearsonCorrelation, Pearson correlation. The SLPF scoring scheme was based on AtPIN.



The resulting final network consisting of the superset of all interactions inferred for all individual nutrient conditions separately and merged via the bridging (jointly occurring) phosphoproteins (Table S1, File S2) was used in all subsequent analyses.

#### Topology characteristics of the reconstructed integrated network

A primary goal of this study was to characterize the regulatory interactions between phosphoproteins that are dynamically activated or inactivated in response to changes of nutrient supply. We addressed this question by using various parameters that are classically used to describe complex interaction networks (Klemm and Bornholdt, 2005). Node degree frequencies in the reconstructed network appear to be relatively constant up to degree values of 4–5 ( $\log_4 = 1.38$ ). Above this value, the degree distribution follows a power law distribution (Figure 4), a commonly found characteristic of molecular interaction networks (Yook et al., 2004; Albert, 2005). Nodes with greater degree values are becoming increasingly unlikely, but the few nodes with high degree levels may serve as central processing hubs (Table S2). In particular, nucleic acid binding proteins, such as transcription factors were found among the proteins with high degree. This class of proteins has previously been found to contain high numbers of clustered phosphorylation sites in so-called phosphorylation hotspots (Riaño-Pachón et al., 2010; Christian et al., 2012).

For the reconstructed integrated network, directionality was determined based on the log ratio of partial variance as introduced in Schäfer and Strimmer (2005) and Opgen-Rhein and Strimmer (2007). Given the applied threshold (see Materials and Methods), directionality was assigned to 343 out of 1067 total interactions. In the following, we further inspected the directed interactions as candidate cause-effect relationships only. We defined proteins with no in-degree but out-degrees of one

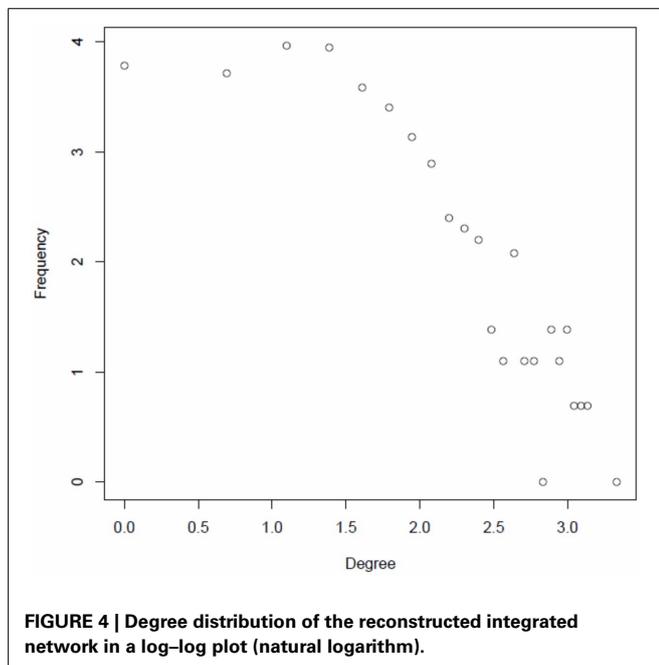
or greater as the “initiation” layer of the network as the signaling cascade is initiated at this level given the network. Proteins with one or greater in- and outgoing edges form the “processing” layer. Proteins with only incoming but no outgoing edges (out degree zero) constitute the “effector” layer as the signal processing ends at those nodes (Figure 5). In the reconstructed network, the numbers of nodes in the initiation/processing/effector layers were identified as 75/98/104, respectively. Thus, the network “fans out” with higher numbers of processing proteins than initiators and still more effectors than processing proteins (Cytoscape file as File S3, list in File S4). Accordingly, the in-degree [median (mean) of 1(1.7)] is significantly lower than the out-degree [median (mean) of 2(2.1)] for processing layer nodes (Mann-Whitney test,  $p$ -value: 0.02). Taken together, the reconstructed network conforms to a topology in which the information appears disseminated to an increasingly broadened set of phosphoproteins.

In the “initiation” layer, we observed an overrepresentation of plasma membrane proteins ( $p = 0.011$ , Fisher Exact test), compared to processing and effector layers. By contrast, the “effector layer” is particularly enriched for proteins with cytosolic location ( $p = 0.0012$ , Fisher Exact Test) and proteins with nuclear location ( $p = 0.0048$ , Fisher Exact Test). Thus, external stimulation induces signaling processes at the plasma membrane that are relayed to cytosolic and nuclear effectors. The enrichment for transcription-related proteins was highest among the effector proteins ( $p = 0.059$ , Fisher Exact Test). Interestingly, also among the effectors, a large number of kinase proteins ( $p = 0.0014$ , Fisher Exact Test) was found, indicating that some of the observed effectors may still be involved in further “processing”, but their targets were not captured within the time window analyzed in this study (30 min).

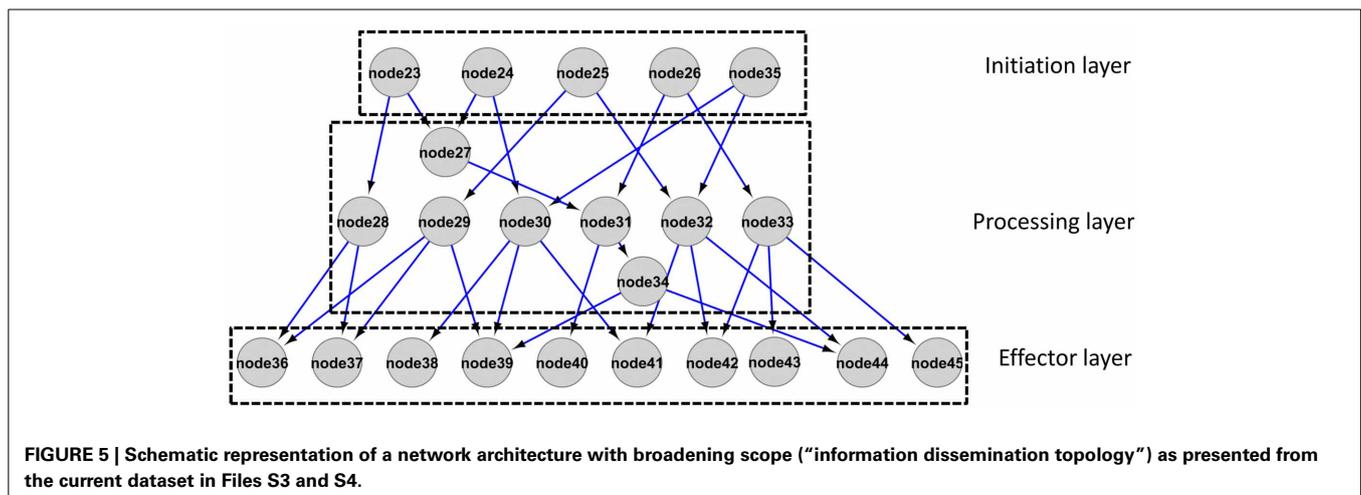
In general, 45% of the proteins of the initiation layer displayed a maximum change point of phosphorylation at 3 min. In contrast, the majority of proteins in the processing and effector layer showed maximum change points in phosphorylation at later time points, such as 10 min (41% for processing, 44% for effectors). Thus, the initiation layer in all treatments largely does refer to the “early” responses particularly involving plasma membrane proteins, while processing and effector layer are enriched in proteins with “late” responses particularly located in the cytosol or nuclear compartments as has already been suggested for nitrogen treatments (Engelsberger and Schulze, 2012).

#### Phosphorylation motifs in the layered network

Based on the layered functional topology of the reconstructed network, we checked for the presence of layer-specific phosphorylation motifs (Figure 6). The initiation layer was found to contain motifs with basic amino acid residues RxxS, which are likely targets of calcium dependent protein kinases. The double-S motif is common among membrane proteins, such as aquaporins and was identified as a common motif of a plant receptor kinase (Wu et al., 2013). The PxxxxS motif is yet uncharacterized. The processing layer was also enriched in the basic RxxS motif again possibly involving members of the CDPK/SnRK family. Interestingly, in the effector layer, we found a significant ( $p$ -value =  $2.32E-05$ , Fisher’s exact test) over-representation of SP motifs, which



**FIGURE 4 | Degree distribution of the reconstructed integrated network in a log–log plot (natural logarithm).**



were shown to be targets for MAP-Kinases. Thus, the identified phosphorylation motifs also reflect the functional layout of the reconstructed network from targets of membrane receptor kinases and CPDKs or PKA in the initiation layer to MAP-Kinase targets in the effector layer. The kinase specificity motifs with acidic amino acids, such as SxxE, QxS and acidic motifs SxxE and SE are likely targets of casein kinase II. These acidic motifs are particularly represented in the processing and effector network layer. These acidic casein kinase II motifs were also particularly present for the protein kinases AT4G08170, AT4G10730, and AT4G38470 (Figure S1), which were the kinases with highest degree in our reconstructed network.

## DISCUSSION

In this study, the dynamics of nutrient-induced protein phosphorylation was investigated. The tested inorganic nutrients ( $\text{NH}_4^+$ ,  $\text{NO}_3^-$ , and  $\text{PO}_4^{3-}$ ) provoked distinctive responses compared to the organic solutes (sucrose, mannitol) with regard to the participating phosphoproteins/peptides and the dynamic change patterns. Based on the time series of peptide phosphorylation levels, an integrated phosphorylation signaling network was reconstructed using a newly designed scoring metric for the selection of the most appropriate network inference method. In general, the topology of the inferred networks corresponds to an information dissemination architecture in which the phosphorylation signal is passed on to an increasing number of processing proteins stratified into an initiation, processing and effector level layer. Specific phosphorylation peptide motifs associated with the three distinct layers were identified indicating the action of layer-specific kinases.

Due to limitations on experimental capacity as well as mass spectrometric instrument time, the time series data used in this study was rather short, but still constituted one of the largest phosphoproteomic time series studied in plants so far. Nonetheless, with only five time points available, reliable network inference is challenging as the statistical power needed to infer interactions between proteins is low. As protein phosphorylation levels were found to change considerably, we conclude that the probed time interval (30 min) was chosen adequately to

capture the initial signaling events even though the presence of protein kinases in the effector layer suggests that the dynamics is still ongoing at the end of the measuring interval. Apart from the limitations with regard to temporal resolution and time interval, the available phosphoprotein set can be considered limited also with regard to phosphoprotein coverage. (Phospho)proteomic data sets collected still suffer from being incomplete due to slow instrument acquisition cycles resulting in lower proteome coverage than is commonly achieved in gene expression profiling using whole-genome microarrays. Particularly, the requirement for quantitative values for all five time points reduced the current data set from about 1000 identified phosphopeptides to the core dataset of approximately 550 peptides with intensity values from label-free quantitation. Out of these, only for 24 peptides full time course data were available from different biological replicates. Nonetheless, we probed whether the underlying data are sufficiently robust by comparing the time profiles of the available repeat measurements for the same peptide, which we expect to be similar, to profile comparisons across different peptides. Indeed, the same-peptide-repeat profiles are much more similar to one another (average Pearson correlation coefficient,  $r_{\text{avg}} = 0.32$ ) than peptide profiles associated with different peptides ( $r_{\text{avg}} = 0.02$ ,  $p_{\text{Mann-Whitney}} = 2.2E - 05$ ). Given the data limitations and technical challenges, the success of network inference attempts critically depends on the selected computational inference method. We performed additional analyses to test the validity of the detected relationships between phosphoproteins.

## EVALUATION OF NETWORK INFERENCE METHODS

We systematically benchmarked a range of different methods by judging their ability to reproduce networks that are consistent with prior biological knowledge. In particular, we used subcellular localization information and consistency of gene ontology annotations to assess the merit of the networks obtained by the various tested methods. We found that relevance-based networks such as linear Pearson correlation and graphical Gaussian method based on regularized partial correlation calculation, and designed initially for large scale gene regulatory network reconstruction, worked best. In a similar study also involving a five-timepoint



**FIGURE 6 | Enriched motifs of phosphopeptides in different layer of the reconstructed network.** Motif-x was used with *p*-value threshold 0.01, the occurrence threshold was set to 10.

phosphoproteomic study in growth-factor stimulated HeLa cells, Pearson correlation was also used successfully in the network reconstruction of SILAC-based quantitative data (Imamura et al., 2010). In general, given the data set limitations, simpler methods; i.e., methods with fewer parameters, may prove superior compared to methods needing large datasets for model and parameter fitting.

In principle, time series data allow detecting cause-effect relationships via time-delay based correlations or using more advanced concepts such as Granger causality (Walther et al., 2010). However, with only five time points available here, such approaches were not applicable. Nonetheless, as those methods bear great potential, efforts should be undertaken to sample as many time points as possible.

We used protein subcellular localization and gene ontology in the network reconstruction method performance assessment. Intuitively, this information could also be used directly during the network inference process. In fact, considering prior information such as known interaction from online databases has been used as additional weighting information in gene regulatory network or signaling network reconstruction efforts integrating prior interaction information in the framework of Bayesian networks (Lo et al., 2012; Pei and Shin, 2012). The assumption that interacting proteins are more prone to interact in the same subcellular compartment (similar functions or related biological pathway) has been used in protein function annotation studies (Van Noort et al., 2003).

Utilizing target-specific motif information as a type of prior information also bears potential for use in phospho-signaling network reconstruction. Based on the assumption that targets of a particular kinase have conserved motifs, kinase-target relationships may be predicted from sequence alone (Yaffe et al., 2001). However, due to the still incomplete understanding of known kinase-substrate motif information for many plant specific kinases, we did not use this information in our approach. Efforts to infer kinase-substrate interaction networks using existing information from large-scale protein-protein interaction studies as prior information may contribute toward expanding the knowledge of kinase-target relationships in the future (Linding et al., 2007; Xue et al., 2008; Song et al., 2012; Newman et al., 2013).

#### SUPPORT FOR THE VALIDITY OF THE RECONSTRUCTED NETWORK

The reconstructed network was compared to available protein-protein interaction databases (Table 4). Eighteen of the predicted edges are also reported in the various PIN-resources as direct physical interactions compared to only seven agreements on average for randomized networks of the same size ( $p < 0.001$ , same number of edges, 1000 randomization runs). In addition, within the experimental phosphopeptide data set, we found higher frequencies of positive correlations between phosphopeptides compared to randomized data sets (Figure S2) suggesting that a systematic signals are contained in the data. While this signal may also originate from overall technical biases, the former result of increased numbers of confirmed correlations via protein-protein interactions supports the view that the detected signal stemmed primarily from the underlying biological processes. The

**Table 4 | Common interactions between the reconstructed network and online PIN databases.**

Partner A	Partner B	Source
AT1G23190	AT5G51820	AtPIN/STRING
AT1G23190	AT1G07110	STRING
AT5G38480	AT1G35580	ANAP/STRING
AT2G27720	AT4G25890	ANAP/STRING
AT5G61780	AT5G07350	STRING
AT1G08420	AT4G03080	AtPIN
AT2G18960	AT4G30190	AtPIN/STRING
AT3G48740	AT2G18960	STRING
AT2G18960	AT1G57990	STRING
AT1G01100	AT2G27710	AtPIN/STRING
AT1G15690	AT1G78900	STRING
AT4G35100	AT2G39010	AtPIN/ANAP/STRING
AT3G09200	AT1G31340	ANAP
AT3G52400	AT1G59870	ANAP/STRING
AT5G45510	AT2G27710	STRING
AT5G37600	AT1G66200	AtPIN
AT3G14350	AT2G39010	STRING
AT1G22280	AT2G40770	STRING

*PIN-databases: AtPIN (Brandão et al., 2009), ANAP (Wang et al., 2012a), and STRING (Franceschini et al., 2013).*

edges identified in our regulatory network as well as in protein-protein interaction networks include interaction in metabolic pathways (PGM—F2KP; 14-3-3—CINV1), ribosomal proteins (RPP1A—RPP2B, RPP2A—RPP3A) as well as the plasma membrane ATPases (AHA1—AHA2). Although these edges are not kinase-substrate pairs, they are likely to be regulated by consecutive events in common pathways. Particularly for ribosomal proteins and the plasma membrane ATPase, regulation by phosphorylation has been described (Fuglsang et al., 2003; Hummel et al., 2012).

Nonetheless, most protein pairs predicted in this study are not reported as direct interactions in the various PIN databases. Aside from acknowledging the obvious possibility that those predicted interactions may be false-positives, we investigated whether the novel, but uncorroborated node pairs may constitute plausible interactions in the sense that they are predicted to occur in a local PIN-environment. We mapped the predicted node pairs onto the integrated PIN (combination of AtPIN, ANAP, and STRING) and computed the shortest path lengths between them. Compared to node pairs drawn at random from the available phosphoprotein set, we found that connected pairs, though not interacting directly, frequently correspond to local interactions as their average shortest path length is significantly smaller (mean = 4.56) than for random pairs (1000 repeats, mean = 4.95, stdev = 0.12,  $p < 0.001$ ). Thus, in many cases, the interactions detected by our time-resolved phosphorylation network inference approach appear to be indirect regulatory interactions for which immediate protein partners are not present in the data set used here. Alternatively, they may correspond to altogether novel interactions as the predicted interactions may be specific for the particular conditions investigated here, which may not have

been captured by other protein–protein interaction studies. For example, several regulatory interactions between protein kinases (e.g., MKK2) and DNA-binding proteins (e.g., RGD3) or transcription factors were identified in the reconstructed network, which were not present in the protein–protein interactions data set. Moreover, the regulatory interactions shown here seem to “shortcut” direct interaction paths.

While we intentionally chose a reconstruction method that best reproduced prior biological information, the actual information itself was not used during the reconstruction process. Furthermore, the protein–protein interactions did not enter directly into the performance assessment metric, but rather the probability of protein–protein encounter based on subcellular localization. Thus, the reported results in support of the validity of the predicted interactions go beyond a circular argument.

### NUTRIENT-INDUCED PHOSPHO-SIGNALING IN ARABIDOPSIS

Even though the detected phosphoprotein sets in the different nutrient treatments overlapped more than expected by chance (Table 2), the generated networks for the individual treatments did not result in common directed interactions between them and could only be merged based on jointly occurring phosphoproteins. Thus, with regard to interactions, the treatments resulted in disjoint sets. Of course, the above mentioned data set limitations may explain this observation. In addition, because the number of possible directed interactions scales with the square of the number of phosphoproteins and, furthermore, the resulting networks are sparse, overlaps between them are less likely than between the proteins sets themselves. If, on the other hand, the result of disjoint interaction sets is taken as the truth, it can be concluded that the different nutrient or osmotic stimuli trigger different signaling cascades. Particularly, it is known that the mere osmotic component (i.e., in mannitol treatment or also sucrose) is primarily perceived through a histidine-kinase sensory system (Urao et al., 1999) which inherently differs from transporter-based sensing systems in nitrogen acquisition (Ho and Tsay, 2010). It is interesting to note that the phosphoproteins commonly found in several different treatment conditions (Table 2) preferentially fall into the processing level layer [55% on average, significantly more than expected by chance ( $p = 1.2E-4$ , binomial distribution processing layer vs. others accounting for layer set sizes as well)]. This supports the network topology with a specialized smaller number of initiators that induce the specific response but feed into common proteins with information processing role. On level of the effectors, again larger specificity could be observed regarding the different treatments.

In biological signaling cascades, an increase of interaction space for downstream kinases has been observed before. An example is the MAP-Kinase cascade, in which the final effector kinase, the MAP-Kinase (MPK) is activated by a MAP-Kinase-Kinase (MKK), which in turn is activated by a MAP-Kinase-Kinase Kinase (MAP3K). In protein array experiments identifying substrates to each of the different levels in the MAP-Kinase cascade, largest numbers of interactions (Popescu et al., 2009) were found with the MPKs (570 substrate candidates) while “incoming” interactions of MPKs with the activating kinases

MKK resulted in far less confirmed interactions (9 MKKs). The MPK signaling pathway is particularly involved in responses to abiotic stress such as salt stress. Thus, the observed links to the MPK cascade could particularly be an osmotic response and was especially observed with the mannitol treatment.

A great diversity was also observed in the plant receptor kinases, many of which feed into soluble kinase signaling pathways (Osakabe et al., 2013). This receptor diversity particularly reflects the plant’s potential to respond to different stimuli with only a small number of receptor/co-receptor pairs being activated by the particular signal and serving as initiators for downstream cascades. Therefore, the topology observed in our reconstructed network is indeed reflected in biological signaling networks with also receptor kinase as members of the initiation layer.

In our analysis, only proteins with phosphoproteomic profiles consisting of five time points were included for network reconstruction. This greatly reduced the data set from about 1000 of identified phosphopeptides in the original studies to only approximately 500 peptides studied here. By imposing directionality, the layered network only included 277 phosphoproteins. Important players in nutrient signaling include the Snf1-related kinase family (SnRK) (Shin et al., 2007) and a central growth regulator TOR (Caldana et al., 2013), which was also shown to interact with nutrient responses. Although members of these kinase families were found to be phosphorylated in response to nitrogen changes (Engelsberger and Schulze, 2012), the data density was not high enough to record full time profiles for these proteins and thus they could not be included in this work. This indicates that key responses do occur at individual time points only and these cases cannot always be clearly distinguished from missing data points due to data-dependent acquisition by mass spectrometry (Schulze and Usadel, 2010). Future challenges lie in efficiently including and biologically weighting single occurrences within the signaling network context.

### CONCLUSION

A primary goal of this study was to characterize the regulatory interactions between phosphoproteins that are dynamically activated or inactivated in response to changes of the nutrient and organic-solute challenges supply. To address this question, we used methods of classic statistical interaction network description and developed a novel scoring metric for inferred networks. Our main findings suggest a hierarchical architecture of regulatory phosphorylation networks with proteins involved in signal initiation, signal processing, and an effector layer. An increasing number of proteins in downstream hierarchical layers of the network were found which is consistent with information processing features typically observed in phosphorylation cascades. Our work demonstrates that even with short time course data sets, network inference methods are applicable for the reconstruction of information processing networks.

### MATERIALS AND METHODS

#### EXPERIMENTAL DATA

Time series phosphoproteomic data were obtained from two published data sets involving sucrose and mannitol treatment

(Niittylä et al., 2007), as well as nitrate and ammonium treatment (Engelsberger and Schulze, 2012), or phosphate treatment. The dataset is publicly available in the PhosPhAt database (Durek et al., 2010). Briefly, in all treatments, the experimental design involved starvation of 14-day old *Arabidopsis* seedling for a particular nutrient and resupply of the respective nutrients for various time periods (3, 5, 10, 30 min) after which samples were taken. The starved seedlings served as controls for that particular nutrient (0 min). Protein extracts were fractionated into plasma membrane and soluble proteins before tryptic digestion and enrichment of phosphopeptides over TiO<sub>2</sub> or IMAC. Phosphopeptides were subsequently analyzed by LC-MS/MS on an Orbitrap mass spectrometer. Label-free quantitation for all experiments was carried out based on ion intensities of phosphopeptides and involved protein correlation profiling to obtain quantitative values for precursor ions without fragmentation spectra.

#### EXPECTED NUMBER OF PHOSPHO-PROTEINS COMMON TO PAIRS OF TREATMENTS

For a given pair of treatments  $T_1$  and  $T_2$ , the expected random number of phosphopeptides observed in both treatments was calculated as  $N_{T_1} \cdot N_{T_2} / N$  assuming independence of the individual observations with  $N$  referring to the total number of *Arabidopsis* phosphoproteins (5237, downloaded on 10.20.2011) contained in the PhosPhAt database (Durek et al., 2010), and  $N_{T_{1/2}}$  referring to the number of phosphoproteins observed in the respective treatment conditions.

#### HIERARCHICAL CLUSTERING OF PHOSPHOPROTEIN SETS

Using a dissimilarity metric (the inverse of the ratio between the actual number and the randomly expected number of phosphoproteins observed jointly), the nutrient treatments were clustered using hierarchical clustering (complete linkage).

#### DATA SET NORMALIZATION

For each treatment, the original intensity values were first log-transformed (natural logarithm). Afterwards, the median value of each time point across all peptides was subtracted from every individual value for a particular peptide and time point.

#### DETECTION OF SIGNIFICANTLY CHANGED PEPTIDE PHOSPHORYLATION LEVELS BETWEEN CONSECUTIVE TIME POINTS

After normalization, phosphorylation of a particular phosphopeptide was considered changed, if the difference  $d$  with  $d = a[t] - a[t - 1]$ ,  $t \geq 2$ , where  $a[t]$  represents the phosphorylation level at time point  $t$ , was outside of a specified threshold range set to  $[-1.71, 1.64]$  for down- and upregulation, respectively. The threshold criteria selection was based on the observed mean and standard deviation of  $d$  across all intervals and across all peptides for all treatments with the interval of no-change determined as  $[\text{mean}(d) - \text{std}(d), \text{mean}(d) + \text{std}(d)]$ . The dynamics of phosphorylation level of a particular P-site was described qualitatively as a succession of “increase,” “decrease,” and “no change” (relative to thresholds defined above) events and referred to as the *change pattern*.

Note that despite phosphorylation being a binary on/off event for any given phosphorylation site, phosphorylation levels were

considered as continuous variables as the measured experimental phosphorylation levels resulted from an average across many protein copies in the sample with most likely not fully coordinated on/off phosphorylation events.

#### HIERARCHICAL CLUSTERING OF CHANGE PATTERNS

Each peptide time series was translated into a succession of four qualitative changes of “increase,” “decrease,” or “no change” based on the thresholds explained above. For each treatment, a frequency vector with length 81 was calculated corresponding to the maximal number of theoretically possible change pattern series ( $3^4 = 81$ ). Each element in the vector represents the proportion of peptides following the corresponding change pattern in a particular treatment. Using the Euclidean distance of the frequency vectors for each pair of treatments, the nutrient treatments were clustered using hierarchical clustering (complete linkage).

#### PROTEIN INTERACTION NETWORKS

PIN data was obtained from the public resources AtPIN (Brandão et al., 2009), as of October 2011, comprising 96,821 interactions between 15,163 *Arabidopsis* proteins including both experimentally verified and predicted interactions. In the assessment of reconstructed interactions, several other online PIN databases [PhosPhAt (Zulawski et al., 2013), ANAP (Wang et al., 2012a), and STRING (Franceschini et al., 2013)] were also used. These databases also collect existing evidence for the potential protein interaction in *Arabidopsis*.

#### NETWORK INFERENCE METHODS

Various methods of network inference methods were applied including correlation based methods (Peng et al., 2009) (Pearson correlation [PearsonCorrelation], partial correlation [FirstOrderPartialCorrelation], mutual information based method [MI] (Sales and Romualdi, 2011), graphical model methods (Schäfer and Strimmer, 2005; Lèbre, 2009) (Dynamic Bayesian network [DBN], graphical Gaussian model [FullOrderPartialCorrelation]). The abbreviation in square brackets are also used in **Table 3**. The R package *parmigene* was used for the mutual information based method. The R packages *space* was used for partial correlation based methods. The R package *GIDBN* was applied for dynamic Bayesian network inference. The R package *GeneNet* was used for the graphical Gaussian model method.

Several of the employed methods require setting threshold parameters, e.g., a minimum correlation coefficient and/or associated  $p$ -value of correlation. Evidently, less strict thresholds result in larger networks. In order to compare methods between each other based on obtained networks of the same size, networks were pruned to a common size of  $3 \cdot N$  ( $N$  is the number of measured phosphopeptides) by eliminating edges and associated nodes in decreasing score order; i.e., high-confidence edges were retained. The size  $3 \cdot N$  was chosen as molecular interaction networks often exhibit an average degree of 2–4 (Werhli, 2012). Networks were first reconstructed for each nutrient treatment condition separately and then merged into a single, final network as the superset of all individual predicted interactions.

## NETWORK INFERENCE METHOD PERFORMANCE ASSESSMENT METRIC

In order to evaluate the accuracy of the reconstructed networks, we devised a novel scoring metric that is based on the subcellular location of proteins. Predicted interactions between proteins were scored by how likely they are based on prior knowledge of physical interactions between proteins from different or the same cellular compartments. For all pairwise physical interactions contained in AtPIN, the subcellular localization of the two interacting proteins as reported in SUBA (Heazlewood et al., 2007) was determined resulting in a comprehensive frequency table of physical interactions between proteins from all compartments termed the subcellular location pair frequency (SLPF, **File S5**). The corresponding subcellular location pair frequency network is shown in **Figure S3**, the width of the edge is proportional to its frequency. For example, it is very likely that two interacting proteins are both from the nucleus as the corresponding edge (arc to itself) has a large width. By contrast, direct physical interactions between proteins located in the nucleus and in the endosome are less likely. Hence, network inference methods can be compared to each other by summing up the SLPF-score computed for all predicted set of interactions dividing subsequently by the number of predicted interactions to account for different network sizes. Better methods will be associated with higher SLPF scores.

In addition to the SLPF-score, we also used the Gene Ontology (GO) semantic similarity method (Wang et al., 2007), which includes not only compartment component, but also molecular functions and biological pathway, as a metric to assess the validity of the reconstructed networks. Here, interactions are scored by how similar the GO annotations of proteins predicted to interact are. Again, scores are summed up for all predicted interactions and divided by their number.

## DIRECTIONALITY OF CORRELATION-BASED INTERACTIONS

After reconstructing the network based on Pearson correlation, we used the method introduced in (Schäfer and Strimmer, 2005; Opgen-Rhein and Strimmer, 2007) to calculate the directionality of each interaction. First, the shrinkage estimate of correlation was computed by function “cor.shrinkage” in R package “corpcor.” Secondly, concentration matrix was calculated as the inverse of the correlation matrix by function “solve” from R package “base.” Finally, the partial variance was obtained as the reciprocal of diagonal element of the concentration matrix, and then the ratio of partial variance was calculated correspondingly. The direction of the arrows points from the node with the larger standardized partial variance (the more “exogenous” variable) to the node with the smaller standardized partial variance (the more “endogenous” variable). The directionality threshold value was set to a log-ratio of partial variances of  $\pm 0.002$ . The threshold was obtained as the average top and bottom 25 percentile of log-ratios across all five treatment conditions.

As networks were reconstructed separately for every treatment, conflicts of directionality assignments are, in principle, possible. However, no directed edge between any two proteins was found common to any pair of treatments.

## NETWORK PROPERTIES

In undirected networks, the degree of a node  $n$  is the number of edges linked to  $n$  (Assenov et al., 2008). The node degree distribution gives the number of nodes with degree  $k$  for  $k = 0, 1, \dots$ . In directed networks, the in-degree of a node  $n$  is the number of incoming edges and the out-degree is the number of outgoing edges.

## FUNCTIONAL TERM ENRICHMENT ANALYSIS

Fisher’s exact test was used in the functional and localization enrichment analysis based on the available Gene Ontology annotation information from TAIR (version 10) (Lamesch et al., 2012), MAPMAN annotation (Thimm et al., 2004) and SUBA subcellular localization information. Specific GO-function terms associated with phosphorylation processes were selected including “common” (=all other categories), kinase, metabolic, transporter, and phosphatase. Selected GO-localization terms were tested to detect preferences with regard to subcellular localization including the terms nucleus, Golgi, mitochondrion, endoplasmic reticulum, cytosol, extracellular, vacuole, plasma membrane, plastid, peroxisome, cytoskeleton, cell plate, endosome.

## SEMANTIC SIMILARITY

The semantic similarity of the gene ontology (GO) terms associated with two genes was calculated based on the “Wang” measure (Wang et al., 2007), which aggregates the semantic contributions of their ancestor terms (including this specific term) in the GO graph. The R package “GOSemSim” was used in the semantic similarity calculations (Yu et al., 2010).

## SUBCELLULAR LOCALIZATION INFORMATION

Subcellular localization information of Arabidopsis proteins was obtained from the SUBA database (Heazlewood et al., 2007; Tanz et al., 2013).

## PHOSPHORYLATION-SITE MOTIF DETECTION

Motif-x was used in the phosphopeptide motif analysis, which is an iterative strategy to build successive motifs through comparison to a dynamic statistical background (Schwartz and Gygi, 2005). The phosphopeptide sequence was extracted with length 13 (the phosphorylation site was in the middle position). The  $p$ -value threshold was set to 0.01 and the occurrence threshold was set to 10 as the motif-x web service setting-up. The whole Arabidopsis proteome was selected as the background reference.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2013.00540/abstract>

**Figure S1 | Phosphopeptide motifs of extracted from peptides associated with proteins which have an absolute correlation coefficient greater than or equal to 0.8 with selected kinases based on the respective phosphorylation level profiles.** The analysis was done for those kinases with at least 10 correlated proteins. Those proteins and their associated peptides were then interpreted as the kinase targets. motif-x (Schwartz and Gygi, 2005) was used with  $p$ -value threshold 0.01, the occurrence threshold was set to 10.

**Figure S2 | Distribution of Pearson correlation coefficients associated with the actual experimental data (black) and randomized data (red lines, 1000 randomizations).**

**Figure S3 | Subcellular location pair frequency network derived from AtPIN and SUBA.** The width of each edge is proportional to its frequency. Cytoscape was used to visualize the network. Predicted interactions can be scored by how likely they are given the expected interaction frequencies displayed here.

**File S1 | List of phosphopeptides including their dynamic change patterns.**

In the “change pattern” column, “N” represents no change at corresponding time point compared to the previous one, “U” and “D” represents up-regulated and down-regulated, respectively.

**File S2 | Cytoscape file of the joint reconstructed Pearson correlation network based on the different nutrient treatments.**

**File S3 | Cytoscape file of directional network in layered hierarchical layout.**

The log ratio of partial variance was set to 0.002 to obtain the directional network for interactions in the network of **File S2** or **Figure S3**.

**File S4 | List of phosphoproteins and their properties in the layered network.** “protein based layer” column shows the layer position for each protein in this directional network.

**File S5 | Interaction frequency for each pair of subcellular localizations.**

The subcellular localizations include nucleus, Golgi, mitochondria, endoplasmic reticulum, cytosol, extracellular, vacuole, plasma membrane, plastid, peroxisome, cytoskeleton, cell plate, endosome, and unclear.

**File S6 | Commonly detected proteins among different pairs of treatments.**

**Table S1 | Reconstructed network summary (number of nodes, number of interactions, and average degree) for each treatment, which was then combined into an integrated “super-” network.**

**Table S2 | List of top 20 nodes (phosphoproteins) with highest degree in the reconstructed network.**

## REFERENCES

- Albert, R. (2005). Scale-free networks in cell biology. *J. Cell Sci.* 118, 4947–4957. doi: 10.1242/jcs.02714
- Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., and Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol.* 8, 1195–1203. doi: 10.1038/ncb1497
- Alexopoulos, L. G., Saez-Rodriguez, J., Cosgrove, B. D., Lauffenburger, D. A., and Sorger, P. K. (2010). Networks inferred from biochemical data reveal profound differences in toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. *Mol. Cell. Proteomics* 9, 1849–1865. doi: 10.1074/mcp.M110.000406
- Assenov, Y., Ramirez, F., Schelhorn, S.-E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics (Oxford, England)* 24, 282–284. doi: 10.1093/bioinformatics/btm554
- Bansal, M., Belcastro, V., Ambesi-Impiombato, A., and Di Bernardo, D. (2007). How to infer gene networks from expression profiles. *Mol. Syst. Biol.* 3:78. doi: 10.1038/msb4100158
- Benschop, J. J., Mohammed, S., O’Flaherty, M., Heck, A. J. R., Slijper, M., and Menke, F. L. H. (2007). Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol. Cell. Proteomics* 6, 1198–1214. doi: 10.1074/mcp.M600429-MCP200
- Birtwistle, M. R., and Kholodenko, B. N. (2009). Endocytosis and signalling: a meeting with mathematics. *Mol. Oncol.* 3, 308–320. doi: 10.1016/j.molonc.2009.05.009
- Blagoev, B., Ong, S.-E., Kratchmarova, I., and Mann, M. (2004). Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* 22, 1139–1145. doi: 10.1038/nbt1005
- Borisov, N. M., Markevich, N. I., Hoek, J. B., and Kholodenko, B. N. (2005). Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys. J.* 89, 951–966. doi: 10.1529/biophysj.105.060533
- Brandão, M. M., Dantas, L. L., and Silva-Filho, M. C. (2009). AtPIN: *Arabidopsis thaliana* protein interaction network. *BMC Bioinform.* 10:454. doi: 10.1186/1471-2105-10-454
- Caldana, C., Li, Y., Leisse, A., Zhang, Y., Bartholomaeus, L., Fernie, A. R., et al. (2013). Systemic analysis of inducible target of rapamycin mutants reveal a general metabolic switch controlling growth in *Arabidopsis thaliana*. *Plant J. Cell. Mol. Biol.* 73, 897–909. doi: 10.1111/tpj.12080
- Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. A., et al. (2009). Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Syst. Biol.* 5:239. doi: 10.1038/msb.2008.74
- Chen, Y., Hoehenwarter, W., and Weckwerth, W. (2010). Comparative analysis of phytohormone-responsive phosphoproteins in *Arabidopsis thaliana* using TiO<sub>2</sub>-phosphopeptide enrichment and mass accuracy precursor alignment. *Plant J.* 63, 1–17. doi: 10.1111/j.1365-313X.2010.04218.x
- Christian, J.-O., Braginets, R., Schulze, W. X., and Walther, D. (2012). Characterization and prediction of protein phosphorylation hotspots in *Arabidopsis thaliana*. *Front. Plant Sci.* 3:207. doi: 10.3389/fpls.2012.00207
- Ciacco, M. F., Wagner, J. P., Chuu, C.-P., Lauffenburger, D. A., and Jones, R. B. (2010). Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat. Methods* 7, 148–155. doi: 10.1038/nmeth.1418
- Cohen, P. (2000). The regulation of protein function by multisite phosphorylation – a 25 year update. *Trends Biochem. Sci.* 25, 596–601. doi: 10.1016/S0968-0004(00)01712-6
- Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Kholodenko, B. N., and Gilles, E. D. (2006). A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinform.* 7:34. doi: 10.1186/1471-2105-7-34
- Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J., Biosystems, P., et al. (2007). “Rule-based modelling of cellular signalling,” in: *Proceedings of the 18th International Conference on Concurrency Theory*, (Lisbon), 17–41.
- Dreze, M., Carvunis, A.-R., Charlotiaux, B., Galli, M., Pevzner, S. J., Tasan, M., et al. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science* 333, 601–607. doi: 10.1126/science.1203877
- Durek, P., Schmidt, R., Heazlewood, J. L., Jones, A., MacLean, D., Nagel, A., et al. (2010). PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. *Nucleic Acids Res.* 38, D828–D834. doi: 10.1093/nar/gkp810
- Ekins, S., Xu, J. J., Alexopoulos, L. G., Saez-Rodriguez, J., and Espelin, C. W. (2008). “High-throughput protein-based technologies and computational models for drug development, efficacy, and toxicity,” in *Drug Efficacy, Safety, and Biologics Discovery: Emerging Technologies and Tools*, eds. S. Ekins and J. J. Xu (Hoboken, NJ: John Wiley & Sons, Inc.), 29–52.
- Engelsberger, W. R., and Schulze, W. X. (2012). Nitrate and ammonium lead to distinct global dynamic phosphorylation patterns when resupplied to nitrogen starved Arabidopsis seedlings. *Plant J.* 69, 978–995. doi: 10.1111/j.1365-313X.2011.04848.x
- Faeder, J. R., Blinov, M. L., and Hlavacek, W. S. (2009). Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol.* 500, 113–167. doi: 10.1007/978-1-59745-525-1\_5
- Feret, J., Danos, V., Krivine, J., Harmer, R., and Fontana, W. (2009). Internal coarse-graining of molecular systems. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6453–6458. doi: 10.1073/pnas.0809908106
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094
- Fuglsang, A. T., Borch, J., Bych, K., Jahn, T. P., Roepstorff, P., and Palmgren, M. G. (2003). The binding site for regulatory 14-3-3 protein in plant plasma membrane H<sup>+</sup>-ATPase: involvement of a region promoting phosphorylation-independent interaction in addition to the phosphorylation-dependent C-terminal end. *J. Biol. Chem.* 278, 42266–42272. doi: 10.1074/jbc.M306707200
- Gat-Viks, I., and Shamir, R. (2007). Refinement and expansion of signaling pathways: the osmotic response network in yeast. *Genom. Res.* 17, 358–367. doi: 10.1101/gr.5750507
- Gaudet, S., Janes, K. A., Albeck, J. G., Pace, E. A., Lauffenburger, D. A., and Sorger, P. K. (2005). A compendium of signals and responses triggered by

- prodeath and prosurvival cytokines. *Mol. Cell. Proteomics* 4, 1569–1590. doi: 10.1074/mcp.M500158-MCP200
- Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., and Millar, A. H. (2007). SUBA: the Arabidopsis subcellular database. *Nucleic Acids Res.* 35, D213–D218. doi: 10.1093/nar/gkl863
- Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E., and Guthke, R. (2009). Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems* 96, 86–103. doi: 10.1016/j.biosystems.2008.12.004
- Hlavacek, W. S., Faeder, J. R., Blinov, M. L., Posner, R. G., Hucka, M., and Fontana, W. (2006). Rules for modeling signal-transduction systems. *Sci. STKE* 2006:re6. doi: 10.1126/stke.3442006re6
- Ho, C.-H., and Tsay, Y.-F. (2010). Nitrate, ammonium, and potassium sensing and signaling. *Curr. Opin. Plant Biol.* 13, 604–610. doi: 10.1016/j.pbi.2010.08.005
- Huang, P. H., Mukasa, A., Bonavia, R., Flynn, R. A., Brewer, Z. E., Cavenee, W. K., et al. (2007). Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12867–12872. doi: 10.1073/pnas.0705158104
- Hummel, M., Cordewener, J. H. G., De Groot, J. C. M., Smeekens, S., America, A. H. P., and Hanson, J. (2012). Dynamic protein composition of *Arabidopsis thaliana* cytosolic ribosomes in response to sucrose feeding as revealed by label free MSE proteomics. *Proteomics* 12, 1024–1038. doi: 10.1002/pmic.201100413
- Imamura, H., Yachie, N., Saito, R., Ishihama, Y., and Tomita, M. (2010). Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data. *BMC Bioinform.* 11:232. doi: 10.1186/1471-2105-11-232
- Janes, K. A., Albeck, J. G., Gaudet, S., Sorger, P. K., Lauffenburger, D. A., and Yaffe, M. B. (2005). A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 310, 1646–1653. doi: 10.1126/science.1116598
- Janes, K. A., and Yaffe, M. B. (2006). Data-driven modelling of signal-transduction networks. *Nat. Rev. Mol. Cell Biol.* 7, 820–828. doi: 10.1038/nrm2041
- Jørgensen, C., Sherman, A., Chen, G. I., Pasculescu, A., Poliakov, A., Hsiung, M., et al. (2009). Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* 326, 1502–1509. doi: 10.1126/science.1176615
- Kholodenko, B. N. (2006). Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* 7, 165–176. doi: 10.1038/nrm1838
- Klemm, K., and Bornholdt, S. (2005). Topology of biological networks and reliability of information processing. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18414–18419. doi: 10.1073/pnas.0509132102
- Kline, K. G., Barrett-Wilt, G. A., and Sussman, M. R. (2010). *In planta* changes in protein phosphorylation induced by the plant hormone abscisic acid. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15986–15991. doi: 10.1073/pnas.1007879107
- Krouk, G., Mirowski, P., LeCun, Y., Shasha, D. E., and Coruzzi, G. M. (2010). Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. *Genom. Biol.* 11:R123. doi: 10.1186/gb-2010-11-12-r123
- Krüger, M., Kratchmarova, I., Blagoev, B., Tseng, Y.-H., Kahn, C. R., and Mann, M. (2008). Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2451–2456. doi: 10.1073/pnas.0711713105
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jørgensen, T. J. D. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 4, 873–886. doi: 10.1074/mcp.T500007-MCP200
- Lèbre, S. (2009). Inferring dynamic genetic networks with low order independencies. *Statist. Appl. Genet. Mol. Biol.* 8:Article 9. doi: 10.2202/1544-6115.1294
- Li, H., Wong, W. S., Zhu, L., Guo, H. W., Ecker, J., and Li, N. (2009). Phosphoproteomic analysis of ethylene-regulated protein phosphorylation in etiolated seedlings of Arabidopsis mutant ein2 using two-dimensional separations coupled with a hybrid quadrupole time-of-flight mass spectrometer. *Proteomics* 9, 1646–1661. doi: 10.1002/pmic.200800420
- Lima, L., Seabra, A., Melo, P., Cullimore, J., and Carvalho, H. (2006). Post-translational regulation of cytosolic glutamine synthetase of *Medicago truncatula*. *J. Exp. Bot.* 57, 2751–2761. doi: 10.1093/jxb/erl036
- Linding, R., Jensen, L. J., Ostheimer, G. J., Van Vugt, M. A. T. M., Jørgensen, C., Miron, I. M., et al. (2007). Systematic discovery of *in vivo* phosphorylation networks. *Cell* 129, 1415–1426. doi: 10.1016/j.cell.2007.05.052
- Lo, K., Raftery, A. E., Dombek, K. M., Zhu, J., Schadt, E. E., Bumgarner, R. E., et al. (2012). Integrating external biological knowledge in the construction of regulatory networks from time-series expression data. *BMC Syst. Biol.* 6:101. doi: 10.1186/1752-0509-6-101
- Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., et al. (2008). Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* 7, 299–307. doi: 10.1074/mcp.M700311-MCP200
- Morcuende, R., Bari, R., Gibon, Y., Zheng, W., Pant, B. D., Bläsing, O., et al. (2007). Genome-wide reprogramming of metabolism and regulatory networks of Arabidopsis in response to phosphorus. *Plant Cell Environ.* 30, 85–112. doi: 10.1111/j.1365-3040.2006.01608.x
- Morris, M. K., Saez-Rodriguez, J., Clarke, D. C., Sorger, P. K., and Lauffenburger, D. A. (2011). Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput. Biol.* 7:e1001099. doi: 10.1371/journal.pcbi.1001099
- Morris, M. K., Saez-Rodriguez, J., Sorger, P. K., and Lauffenburger, D. A. (2010). Logic-based models for the analysis of cell signaling networks. *Biochemistry* 49, 3216–3224. doi: 10.1021/bi902202q
- Nelander, S., Wang, W., Nilsson, B., She, Q.-B., Pratilas, C., Rosen, N., et al. (2008). Models from experiments: combinatorial drug perturbations of cancer cells. *Mol. Syst. Biol.* 4:216. doi: 10.1038/msb.2008.53
- Newman, R. H., Hu, J., Rho, H.-S., Xie, Z., Woodard, C., Neiswinger, J., et al. (2013). Construction of human activity-based phosphorylation networks. *Mol. Syst. Biol.* 9:655. doi: 10.1038/msb.2013.12
- Niittylä, T., Fuhsang, A. T., Palmgren, M. G., Frommer, W. B., and Schulze, W. X. (2007). Temporal analysis of sucrose-induced phosphorylation changes in plasma membrane proteins of Arabidopsis. *Mol. Cell. Proteomics* 6, 1711–1726. doi: 10.1074/mcp.M700164-MCP200
- Van Noort, V., Snel, B., and Huynen, M. A. (2003). Predicting gene function by conserved co-expression. *Trends Genet.* 19, 238–242. doi: 10.1016/S0168-9525(03)00056-8
- Oliveira, I. C., and Coruzzi, G. M. (1999). Carbon and amino acids reciprocally modulate the expression of glutamine synthetase in Arabidopsis. *Plant Physiol.* 121, 301–310. doi: 10.1104/pp.121.1.301
- Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., et al. (2006). Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635–648. doi: 10.1016/j.cell.2006.09.026
- Opgen-Rhein, R., and Strimmer, K. (2007). From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.* 1:37. doi: 10.1186/1752-0509-1-37
- Osakabe, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L.-S. P. (2013). Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. *J. Exp. Bot.* 64, 445–458. doi: 10.1093/jxb/ers354
- Osuna, D., Usadel, B., Morcuende, R., Gibon, Y., Bläsing, O. E., Höhne, M., et al. (2007). Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J.* 49, 463–491. doi: 10.1111/j.1365-313X.2006.02979.x
- Pawson, T. (1997). Signaling through scaffold, anchoring, and adaptor proteins. *Science* 278, 2075–2080. doi: 10.1126/science.278.5346.2075
- Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science* 300, 445–452. doi: 10.1126/science.1083653
- Pawson, T., and Taylor, L. (2009). Protein phosphorylation goes negative. *Mol. Cell* 34, 139–140. doi: 10.1016/j.molcel.2009.04.007
- Pei, B., and Shin, D.-G. (2012). Reconstruction of biological networks by incorporating prior knowledge into bayesian network models. *J. Comput. Biol.* 19, 1324–1334. doi: 10.1089/cmb.2011.0194
- Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Am. Statist. Assoc.* 104, 735–746. doi: 10.1198/jasa.2009.0126
- Piques, M., Schulze, W. X., Höhne, M., Usadel, B., Gibon, Y., Rohwer, J., et al. (2009). Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Mol. Syst. Biol.* 5:314. doi: 10.1038/msb.2009.68

- Popescu, S. C., Popescu, G. V., Bachan, S., Zhang, Z., Gerstein, M., Snyder, M., et al. (2009). MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Develop.* 23, 80–92. doi: 10.1101/gad.1740009
- Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., et al. (2009). Large-scale *Arabidopsis phosphoproteome* profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol.* 150, 889–903. doi: 10.1104/pp.109.138677
- Riño-Pachón, D. M., Kleessen, S., Neigenfind, J., Durek, P., Weber, E., Engelsberger, W. R., et al. (2010). Proteome-wide survey of phosphorylation patterns affected by nuclear DNA polymorphisms in *Arabidopsis thaliana*. *BMC Genom.* 11:411. doi: 10.1186/1471-2164-11-411
- Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308, 523–539. doi: 10.1126/science.1105809
- Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klant, S., et al. (2009). Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* 5:331. doi: 10.1038/msb.2009.87
- Sales, G., and Romualdi, C. (2011). Parmigene - a parallel R package for mutual information estimation and gene network reconstruction. *Bioinformatics* 27, 1876–1877. doi: 10.1093/bioinformatics/btr274
- Santos, S. D. M., Verveer, P. J., and Bastiaens, P. I. H. (2007). Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat. Cell Biol.* 9, 324–330. doi: 10.1038/ncb1543
- Schäfer, J., and Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21, 754–764. doi: 10.1093/bioinformatics/bti062
- Scheible, W. R., Gonzalez-Fontes, A., Lauerer, M., Muller-Rober, B., Caboche, M., and Stitt, M. (1997). Nitrate acts as a signal to induce organic acid metabolism and repress starch metabolism in tobacco. *Plant Cell* 9, 783–798. doi: 10.1105/tpc.9.5.783
- Schlatter, R., Philippi, N., Wangorsch, G., Pick, R., Sawodny, O., Borner, C., et al. (2011). Integration of Boolean models exemplified on hepatocyte signal transduction. *Briefings Bioinform.* 13, 365–376. doi: 10.1093/bib/bbr065
- Schulze, W. X. (2010). Proteomics approaches to understand protein phosphorylation in pathway modulation. *Curr. Opin. Plant Biol.* 13, 280–287. doi: 10.1016/j.pbi.2009.12.008
- Schulze, W. X., and Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. *Annu. Rev. Plant Biol.* 61, 491–516. doi: 10.1146/annurev-arplant-042809-112132
- Schwartz, D., and Gygi, S. P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* 23, 1391–1398. doi: 10.1038/nbt1146
- Shin, R., Alvarez, S., Burch, A. Y., Jez, J. M., and Schachtman, D. P. (2007). Phosphoproteomic identification of targets of the *Arabidopsis* sucrose nonfermenting-like kinase SnRK2.8 reveals a connection to metabolic processes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6460–6465. doi: 10.1073/pnas.0610208104
- Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431–432. doi: 10.1093/bioinformatics/btq675
- Sneddon, M. W., Faeder, J. R., and Emonet, T. (2011). Efficient modeling, simulation and coarse-graining of biological complexity with Nfsim. *Nat. Methods* 8, 177–183. doi: 10.1038/nmeth.1546
- Song, C., Ye, M., Liu, Z., Cheng, H., Jiang, X., Han, G., et al. (2012). Systematic analysis of protein phosphorylation networks from phosphoproteomic data. *Mol. Cell. Proteomics* 11, 1070–1083. doi: 10.1074/mcp.M111.012625
- Sugiyama, N., Masuda, T., Shinoda, K., Nakamura, A., Tomita, M., and Ishihama, Y. (2007). Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol. Cell. Proteomics* 6, 1103–1109. doi: 10.1074/mcp.T600060-MCP200
- Tanz, S. K., Castleden, I., Hooper, C. M., Vacher, M., Small, I., and Millar, H. A. (2013). SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Res.* 41, D1185–D1191. doi: 10.1093/nar/gks1151
- Terfve, C., and Saez-Rodriguez, J. (2012). Modeling signaling networks using high-throughput phospho-proteomics. *Adv. Exp. Med. Biol.* 736, 19–57. doi: 10.1007/978-1-4419-7210-1\_2
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-313X.2004.02016.x
- Urao, T., Yakubov, B., Satoh, R., Yamaguchi-Shinozaki, K., Seki, M., Hirayama, T., et al. (1999). A transmembrane hybrid-type histidine kinase in *Arabidopsis* functions as an osmosensor. *Plant Cell* 11, 1743–1754. doi: 10.1105/tpc.11.9.1743
- Walther, D., Strassburg, K., Durek, P., and Kopka, J. (2010). Metabolic pathway relationships revealed by an integrative analysis of the transcriptional and metabolic temperature stress-response dynamics in yeast. *Omics* 14, 261–274. doi: 10.1089/omi.2010.0010
- Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., and Chen, C.-F. (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics* 23, 1274–1281. doi: 10.1093/bioinformatics/btm087
- Wang, C., Marshall, A., Zhang, D.-B., and Wilson, Z. A. (2012a). ANAP: an integrated knowledge base for *Arabidopsis* protein interaction network analysis. *Plant Physiol.* 158, 1523–1533. doi: 10.1104/pp.111.192203
- Wang, Y.-Y., Hsu, P.-K., and Tsay, Y.-F. (2012b). Uptake, allocation and signaling of nitrate. *Trends Plant Sci.* 17, 458–467. doi: 10.1016/j.tplants.2012.08.007
- Watterson, S., Marshall, S., and Ghazal, P. (2008). Logic models of pathway biology. *Drug Disc. Today* 13, 447–456. doi: 10.1016/j.drudis.2008.03.019
- Werhli, A. V. (2012). Comparing the reconstruction of regulatory pathways with distinct Bayesian networks inference methods. *BMC Genom.* 13:S2. doi: 10.1186/1471-2164-13-S5-S2
- Wu, X. N., Sanchez-Rodriguez, C., Pertl-Obermeyer, H., Obermeyer, G., and Schulze, W. X. (2013). Sucrose-induced receptor kinase SIRK1 regulates plasma membrane aquaporins in *Arabidopsis*. *Mol. Cell. Proteomics* 12, 2856–2873. doi: 10.1074/mcp.M113.029579
- Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., and Yao, X. (2008). GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol. Cell. Proteomics* 7, 1598–1608. doi: 10.1074/mcp.M700574-MCP200
- Yaffe, M. B., Lepar, G. G., Lai, J., Obata, T., Volinia, S., and Cantley, L. C. (2001). A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.* 19, 348–353. doi: 10.1038/86737
- Yook, S.-H., Oltvai, Z. N., and Barabási, A.-L. (2004). Functional and topological characterization of protein interaction networks. *Proteomics* 4, 928–942. doi: 10.1002/pmic.200300636
- Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26, 976–978. doi: 10.1093/bioinformatics/btq064
- Zulawski, M., Braginets, R., and Schulze, W. X. (2013). PhosPhAt goes kinases—searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. *Nucleic Acids Res.* 41, D1176–D1184. doi: 10.1093/nar/gks1081

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 September 2013; accepted: 12 December 2013; published online: 24 December 2013.

Citation: Duan G, Walther D and Schulze WX (2013) Reconstruction and analysis of nutrient-induced phosphorylation networks in *Arabidopsis thaliana*. *Front. Plant Sci.* 4:540. doi: 10.3389/fpls.2013.00540

This article was submitted to *Plant Systems Biology*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2013 Duan, Walther and Schulze. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.