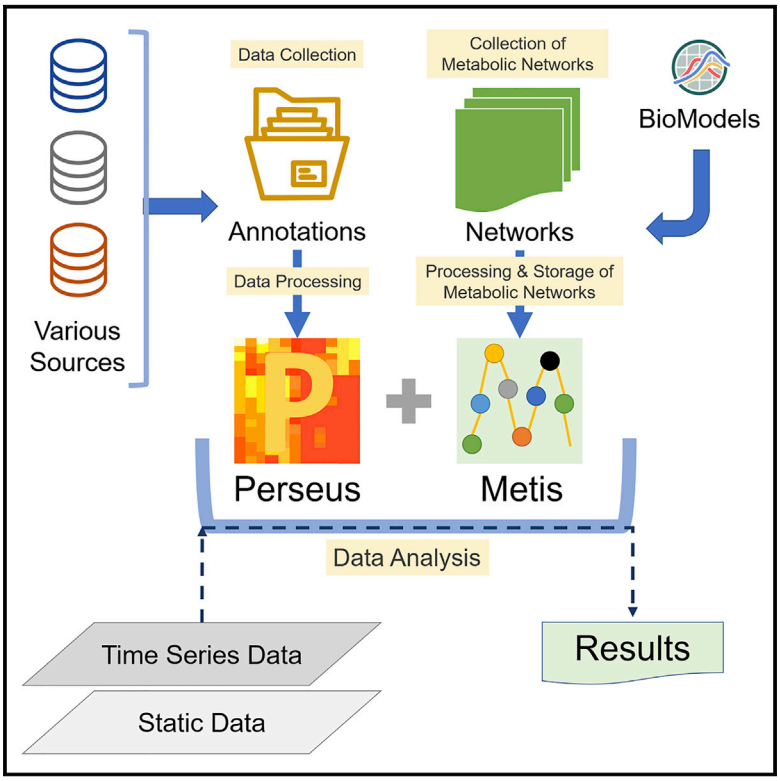
Cell Reports
Methods

Article

Perseus plugin "Metis" for metabolic-pathway-centered quantitative multi-omics data analysis for static and time-series experimental designs

Graphical abstract



Authors

Hamid Hamzeiy, Daniela Ferretti, Maria S. Robles, Jürgen Cox

Correspondence

charo.robles@med.uni-muenchen.de (M.S.R.),
cox@biochem.mpg.de (J.C.)

In brief

Hamzey et al. introduce the Metis plugin of the Perseus software for the analysis of multi-omics data. Metis integrates quantitative data through genome-scale metabolic-pathway reconstructions. Circadian time-series data of the metabolome and phosphoproteome are integrated to reveal examples of modulation of enzymatic activity by phosphorylation.

Highlights

- Perseus software plugin for multi-omics data analysis based on metabolic pathways
- Genome-scale pathway reconstructions link metabolite, transcript, proteome, and PTM data
- Circadian co-cycling of mouse liver phosphoproteome and metabolome is studied
- Contribution of phosphorylation to modulation of enzyme activity is indicated

Hamzeiy et al., 2022, Cell Reports Methods 2, 100198
April 25, 2022
https://doi.org/10.1016/j.crmeth.2022.100198

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Perseus plugin "Metis" for metabolic-pathway-centered quantitative multi-omics data analysis for static and time-series experimental designs

Hamid Hamzeiy,[1] Daniela Ferretti,[1] Maria S. Robles,[2,*] and Jürgen Cox[1,3,4,*]
[1]Computational Systems Biochemistry Research Group, Max Planck Institute of Biochemistry, Martinsried, Germany
[2]Institute of Medical Psychology, Faculty of Medicine, LMU, Munich, Germany
[3]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway
[4]Lead contact
*Correspondence: charo.robles@med.uni-muenchen.de (M.S.R.), cox@biochem.mpg.de (J.C.)
https://doi.org/10.1016/j.crmeth.2022.100198

MOTIVATION Studying two or more types of biomolecules simultaneously in omics studies is of great benefit since it can reveal information that is not apparent when each of the omics dimensions is considered separately. For instance, studying transcriptome and proteome may reveal nodes of post-transcriptional regulation (Buccitelli and Selbach, 2020; Cox and Mann, 2012) that are not apparent in the transcriptomic or proteomic data alone. Another important example is expression quantitative trait loci (Cheung et al., 2005; Morley et al., 2004), in which genetic association is correlated with gene expression to shed light on the relationship between traits and expression-driven cellular processes. The combined analysis of multiple omics dimensions is challenging for multiple reasons. First of all, the quantitative measurements in each technology separately have to be of sufficiently high quality before correlations with other domains can make sense. Further obstacles for multi-omics analysis are limited dynamic range and the resulting missing-value problem inherent to most omics technologies. Furthermore, a statistical challenge arises when performing all-against-all comparisons of variables in one technology with variables in another technology. The number of statistical tests for pairwise correlations explodes and, therefore, either a large number of false positives is created when working with p value thresholds or potentially meaningful truly positive signals are lost due to the necessity of stringent false-discovery-rate control.

SUMMARY

We introduce Metis, a new plugin for the Perseus software aimed at analyzing quantitative multi-omics data based on metabolic pathways. Data from different omics types are connected through reactions of a genome-scale metabolic-pathway reconstruction. Metabolite concentrations connect through the reactants, while transcript, protein, and protein post-translational modification (PTM) data are associated through the enzymes catalyzing the reactions. Supported experimental designs include static comparative studies and time-series data. As an example for the latter, we combine circadian mouse liver multi-omics data and study the contribution of cycles of phosphoproteome and metabolome to enzyme activity regulation. Our analysis resulted in 52 pairs of cycling phosphosites and metabolites connected through a reaction. The time lags between phosphorylation and metabolite peak show non-uniform behavior, indicating a major contribution of phosphorylation in the modulation of enzymatic activity.

INTRODUCTION

Here, we introduce a software solution to this problem for the cases in which a multi-omics analysis involves untargeted metabolome data in combination with one or more other omics technologies, which we connect through the reactions of a metabolic pathway. For this purpose, we developed Metis, which is a plugin for the Perseus software (Tyanova et al., 2016) and which we describe in this work. Perseus is a comprehensive platform for omics data analysis, which was developed with a user in mind who is a life-science researcher but does not necessarily hold a degree in bioinformatics. Hence, we expect to enable a large user base with this type of comparative multi-omics analysis in contrast to other

software tools that are targeted at programmers or bioinformatics specialists.

While Metis can be applied to any kind of experimental design, we focus here on an example with time-series data highlighting an application to circadian multi-omics integration. Circadian rhythms are endogenous and self-sustainable oscillators, present in most living organisms, that drive daily cycles of molecular and metabolic processes (Finger et al., 2020). The molecular mechanism of the clock, built on transcriptional and translational feedback loops, regulates the expression of ~20% of the genes in any given tissue in mammals. Additionally, post-transcriptional and -translational mechanisms are reported to play an essential role in circadian regulation of metabolism (Robles et al., 2014, 2017). Metis allows the investigation of cross correlations between quantitative changes of a metabolic enzyme at different molecular level, such as transcript, protein, and phosphorylation statuses, and the abundance changes of the products and reactants of its catalyzed reactions, aiding to pinpoint key regulatory enzymatic mechanisms. Regulatory nodes could be modulated by phosphorylation-dependent enzyme activity but also by enzyme expression changes at the protein and/or transcript level, which can be distinguished by integrating the proteome and transcriptome in the post-translational modification (PTM) data analysis. The integration of diverse temporal dynamic omics datasets together with rhythmic metabolite profiles from mouse liver uncover phosphorylation as a major enzymatic regulatory mode. Rhythmic phosphorylation of metabolic enzymes regulates its temporal activity and thus metabolic reactions across the day.

RESULTS

Perseus software and Metis plugins

The Perseus software (Tyanova et al., 2016) is a comprehensive framework for high-dimensional omics data analysis with a focus on intuitive usability by interdisciplinary users. Through its plugin architecture, it is extensible by writing code for workflow activities in multiple programming languages, like C#, R, and Python (Yu et al., 2020). Besides statistical analysis on data matrices, the study of networks is supported as well (Rudolph and Cox, 2019). For instance, the PHOTON plugin (Rudolph et al., 2016) can be used to analyze phosphoproteomics data in the context of protein-protein-interaction networks with the aim of reconstructing kinase activities (Brüning et al., 2019).

Here, we extend the network capabilities of Perseus to the specific requirements of metabolic pathways with the Metis toolbox. We use genome-scale networks of metabolic reactions to interconnect data from different omics dimensions (Figure 1). We take reactions, reactants, and enzymes as nodes of the network, while edges connect the reactants and enzymes to the reactions they are taking part in. The enzyme nodes can incorporate multiple types of quantitative omics data such as proteomics, phosphoproteomics, and transcriptomics data. Moreover, nodes can associate to multiple quantitative datasets of different experimental designs, comprising, for instance, sample group comparisons or time-series data. This allows us to compare metabolic reactions across diverse datasets spanning many conditions and containing temporal data, thus providing relevant functional biological information. While we focus, in this work, on the analysis of multi-omics time-series data, Metis is more generic in terms of experimental designs and is also capable of analyzing non-time-series data (Figure S1).

The alternative to this network-based analysis of multiple omics dimensions would be an approach based on all pairwise comparisons between the molecules in the omics datasets without filtering through a network. However, the connection of omics features through the reaction network is crucial to the analysis for the reason of statistical significance of comparisons. To illustrate this, we provide the following example: assume a comparison of untargeted metabolomics data with 1,000 compounds profiled over a time series with 10,000 phosphosites from samples of the same time series. All pairwise comparisons between phosphosites and metabolites would amount to 10,000,000 pairs. One would then perform one statistical test per pair, for instance, to check if the Pearson or Spearman correlation is significantly different from zero. Considering in additional time-lagged correlations would further increase the number of tests on the order of 10-fold. Using a moderate p value threshold would result in many false positives with such a high number of hypotheses tested simultaneously. For instance, a p value threshold of 0.01 would result in on the order of 1,000,000 false significant calls (including an order of magnitude factor 10 for trying different time lags in the correlation calculations). The proper alternative to this would be false discovery rate (FDR) control, for instance with randomizations for generating the null-hypothesis distribution or by applying the Benjamini-Hochberg correction. This would, however, likely not call out any of the tests as significant due to the large background of noisy comparisons. Indeed, for the example presented later, we would have to compare 7,986 profiles of phosphorylation sites with 224 metabolite profiles. Calculating all time-lagged cross correlations with, for instance, five different time lags, including a time lag of zero, results in almost nine million p value calculations that, after Benjamini-Hochberg correction, do not result in any of them being significant. Hence, the comparison of multiple omics through a network that provides *a priori* knowledge about the relationship between the omics levels is crucial for their statistical analysis. In our circadian multi-omics time-series analysis, we implement as a further simplification that we do not use time-lagged correlations but instead use the results of separate cycling analysis per omics in a multi-omics network filtering analysis. This is necessary due to the variation introduced by the heterogeneity of datasets produced in different labs and the general limitations of quantitative accuracy in the available datasets. Since Metis is a general framework for the metabolic-network-based analysis of multi-omics data, the replacement of co-cycling filtering by time-lagged correlations is viable as soon as more accurate and homogeneously acquired data is available.

Rhythmicity estimation of multi-omics circadian time-series datasets

We decided to perform an integrative multi-omics analysis of publicly available datasets assaying *in vivo* circadian dynamics in mouse liver. In order to achieve this, we obtained and re-analyzed the most comprehensive published omics studies in

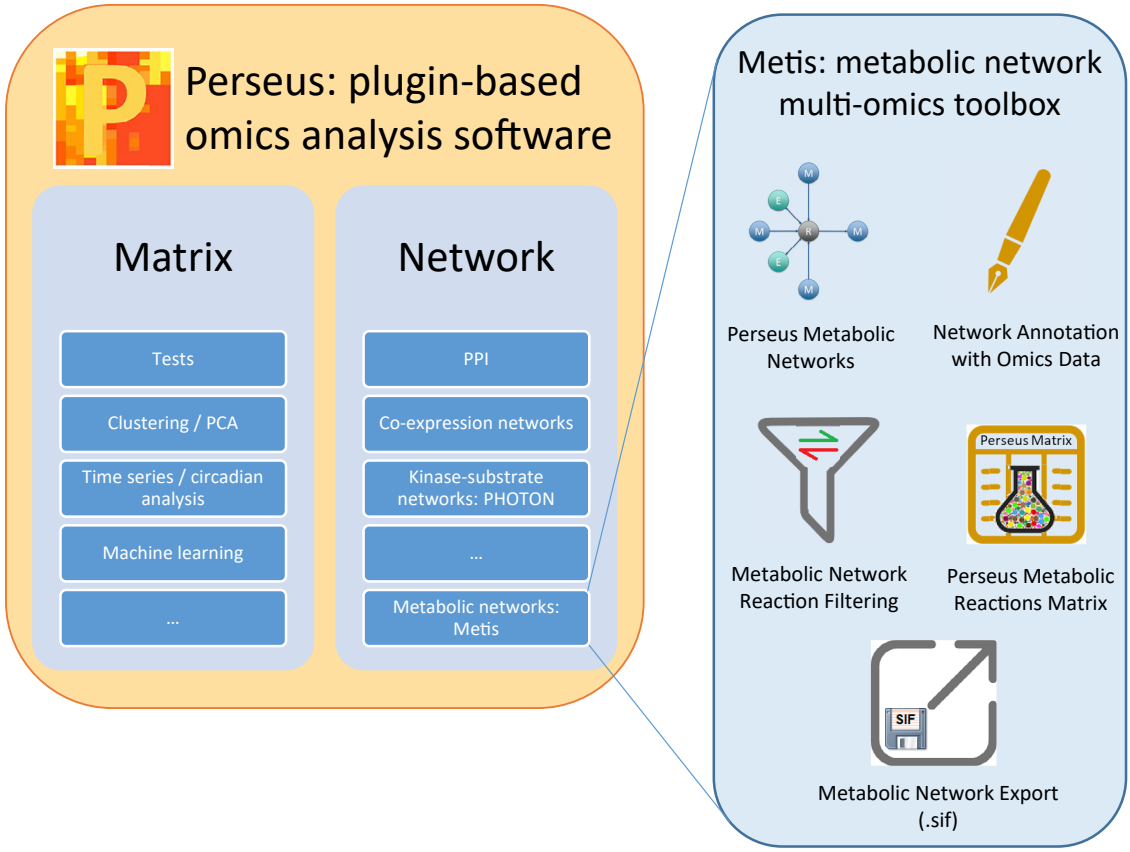


Figure 1. Perseus framework and Metis toolbox

Perseus is a plugin-based software for the analysis of omics data. It supports two basic data types: matrices and networks. The former usually carries relative bimolecular concentration data, and a multitude of activities exist in Perseus to process them. The generic network data structure consists of annotated nodes and edges to accommodate diverse biological networks. Standard plugins contain activities for the creation, import processing, and analysis of, for instance, protein-protein-interaction networks or networks consisting of kinase-substrate relations. The Metis toolbox extends the Perseus network framework to metabolic pathways consisting of reactions, connecting metabolites that are consumed or created with the catalyzing enzymes. Annotation of these networks connects metabolomics matrix data with matrices related to enzymes, which can be proteomics, transcriptomics, or phosphoproteomics data. Network nodes and edges can then be filtered with simultaneous criteria on multiple omics types, resulting in a metabolic reaction matrix containing the results of interest. Finally, results can be exported in .sif format, for instance, for visualization in Cytoscape.

transcriptomics (Hughes et al., 2009), proteomics (Robles et al., 2014), phosphoproteomics (Robles et al., 2017), metabolomics (Krishnaiah et al., 2017), and lipidomics (Adamovich et al., 2014) (Figure 2; Table 1). An important selection criterion was to choose studies done using the same animal housing conditions. In all studies, mice were entrained to light-dark cycles prior to being released to constant darkness, allowing us to use the same periodicity analysis in all datasets. Consequently, we used 23.6 h as the period length, since this is the approximate free-running period of mice driven by the internal clock. Sampling resolution differed among those studies, varying from 1 to 4 h, and we kept the original time resolution of each dataset for the integrative re-analysis. Using the Perseus cycling analysis package (STAR Methods), we analyzed rhythmicity in all omics datasets individually by fitting the log-transformed data to a cosine curve with a period of 23.6 h and calculating the FDR using 1,000 randomizations to simulate the null hypothesis of no-cycling behavior. To avoid discrepancies with the findings in the published datasets, we used the same FDR cut offs as in the original publications. In addition, we used the phosphoproteomics dataset to predict kinase activity using the PHOTON method (Rudolph et al., 2016). The Perseus session for the cycling analysis plus the respective software version are provided as supplemental information.

Cycling biomolecules

The resulting cycling analysis of the five omics datasets plus the predicted kinases with daily patterns of activity is represented in Figure 3. The total number of cycling molecules has to be interpreted with caution, since it is biased by the depths of the respective technologies. The fraction of cycling molecules relative to all molecules quantified by the technology (Figure 3A) is more meaningful but still not free from biases. For instance, the ability to detect statistically significant cycling profiles strongly depends on the quantitative precision of the technology, the number of time points used per cycle, and the total length of the time series in relation to the period length. Furthermore, within datasets, molecules that are close to the detection limit

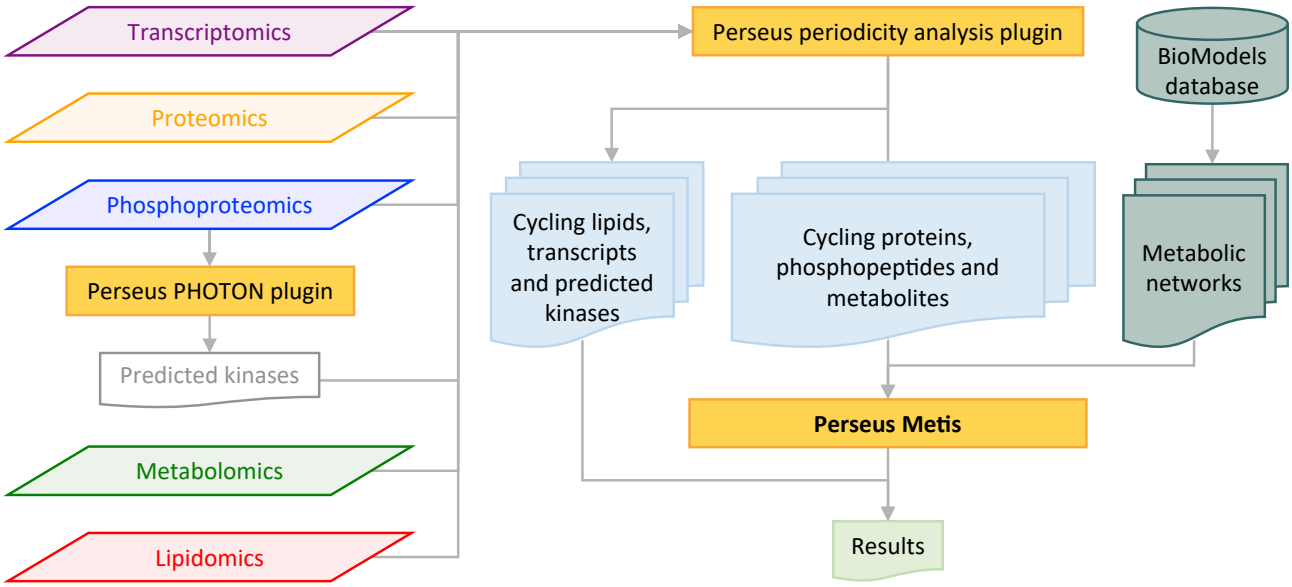


Figure 2. Schematic overview of the data-analysis workflow with time-series circadian data

Datasets used from five different studies of circadian rhythms in mice liver serve as inputs along with the results of the kinase-activity prediction using the PHOTON plugin within the Perseus software package. These were independently re-analyzed using the periodicity analysis toolkit in the Perseus software. The results were then merged with the reconstructed mouse metabolic network from the BioModels database using the network-analysis module of Perseus.

tend to be more difficult to be considered as cycling compared with highly abundant molecules.

The profiles of all cycling biomolecules are shown as heatmaps in Figure 3B, which are sorted in the vertical direction by their acrophase. Out of all biomolecules, metabolites show, with 58%, the largest fraction of cycling molecules. Next frequent are transcripts, of which almost one-third show rhythms in abundance with acrophases uniformly distributed across the day, as described in the original publication and in other pioneering studies done with wider-spaced time points (Panda et al., 2002; Storch et al., 2002; Ueda et al., 2002). In contrast, mouse liver proteins have a smaller cycling fraction, 6% of the total, sharply peaking at two main clusters, one during the day and a second one in the middle of the night (Figure 3B). The latter cluster is due to the induction of protein translation in response to an increase in energy levels due to feeding, occurring during the night in mice as nocturnal animals. It was found (Robles et al., 2014) that when filtering the transcripts to those for which the corresponding protein is cycling, phase relations between peaking proteins and transcripts are on average compatible with the expected time lag between transcription and translation but with strong variations between individual transcript-protein pairs.

Protein function is often regulated by PTMs rather than, or in addition to, changes in protein levels. This is the case for many proteins involved in temporally regulated signaling pathways in the liver (Robles et al., 2017). Thus, it is not surprising that 26% of the phosphorylation sites, corresponding to more than 40% of liver proteins, display daily rhythms almost completely independent of protein-abundance cycles. Similar to rhythmic proteins, cycling phosphorylation showed two distinct clusters of acrophases, with the majority of peaks occurring during the day or resting phases and slightly earlier than the protein cluster. The extensive regulation of phosphorylation in mouse liver implies temporal control of kinase activity. Kinase activity is, in addition to protein levels, strongly regulated by phosphorylation and often by autophosphorylation. Thus, while among all cycling proteins in liver, there are three kinases with rhythmic changes, 55 kinases display cycles of phosphorylation abundance. It is, however, very challenging to infer kinase activity by using phosphorylation patterns on the kinases since the majority of phosphorylation sites are of unknown function (Needham et al., 2019). Taking advantage of curated kinase-substrate relationships (Hornbeck et al., 2015), we were previously able to infer a number of kinases whose activity oscillates across the day (Robles et al., 2017). Since this prediction method is biased toward well-known kinases, we here use an alternative method to predict cycles of kinase activity: the PHOTON algorithm (Rudolph et al., 2016), which is based on the statistical analysis of protein-protein-interaction networks. Applying it to the phosphoproteome data, we were able to predict 33 distinct kinases with changes in their activities across the day, corresponding to 20% of all kinases in the PHOTON analysis. Interestingly, predicted kinase activities are enriched in two temporal regions slightly preceding the phosphorylation clusters on average, indicative of a time lag between peak kinase activity and maximum substrate phosphorylation. Overall, our data show that kinetics of molecular reactions, such as phosphorylation, can be studied using large-scale time-series data.

Circadian clocks and metabolism crosstalk bidirectionally. While tissue clocks regulate local metabolism, the metabolic state feeds back to the molecular clock (Brown, 2016). Accordingly, in mouse liver, which is one of the most studied organs, a large proportion of metabolites have been described to display rhythms across the day. Perseus cycling analysis of the

Table 1. Datasets used in this study

Paper	Hughes et al. (2009)	Robles et al. (2014)	Robles et al. (2017)	Krishnaiah et al. (2017)	Adamovich et al. (2014)
Dataset	transcriptomics	proteomics	phosphoproteomics	metabolomics	lipidomics
Number of identified molecules	18,647	3,132	7,986	224	159
Number of cycling molecules	5,989	186	2,066	131	11
Total duration of sampling (hours)	48	48	48	48	20
Number of time points measured	48 (every hour)	16 (every 3 h)	16 (every 3 h)	48 (every hour)	6 (every 4 h)
Replicates per time point	1	3	3	4	4
Cycling q value threshold	0.05	0.33	0.1	0.05	0.05

Details of the five omics datasets of circadian mouse liver entrained in day-night cycles and then free running from time point 0. Several details of the time-series acquisition, such as the total acquisition time, the sampling rate, and the number of replicates per time point, vary. Cycling q values differed between the analyses performed in the respective publications. In order to keep consistency with previous work, we applied the cycling q value that was used in each publication.

metabolomics data resulted in more than 50% of metabolites with daily cycles, similar to what was reported in the original study (Krishnaiah et al., 2017). Rhythmic metabolites peak at diverse times of the day, with many of them during the inactive phase (Figure 3). Similarly, Perseus cycling analysis of the lipidomics data yielded cycling lipids, 7% of the total, peaking predominantly during the day as previously reported (Adamovich et al., 2014).

Reaction-based multi-omics filtering

Metabolic networks are represented within Metis with three different node types, namely reaction, enzyme, and metabolite (Figure 4). Metabolite nodes are connected to the reactions they participate in and are classified as a substrate or a product of the reaction in question. Since enzymes are not consumed or produced via the reaction but only catalyze it, these nodes are connected to the reaction nodes in an undirected manner. All node types can be annotated within the nodes table of the network in Perseus with both qualitative and quantitative information, e.g., in the case of a reaction, this can be its reversibility or rate. Edges can also be annotated with various qualitative and quantitative information. When filtering for reactions of interest, Perseus can retain all reaction nodes where a condition is either true or false and/or a certain threshold is applied to the numerical annotation at the metabolite or enzyme nodes of the metabolic network. In other words, all nodes that meet the conditions of the filter plus the reaction nodes directly connected to them can either be retained or removed in order to reach the desired network for further analysis in the next processing step. The user also has the option to apply several filters of a certain type on different properties of the nodes and edges in union or as an intersect, depending on the nature of the question. Here, we apply a filter on the q values of the periodicity analysis of each of the three omics dimensions, namely proteomics ($q \geq 0.33$), phosphoproteomics ($q < 0.1$), and metabolomics ($q < 0.05$), to retain all reactions that have nodes with our criteria for further analysis. Following this filtering, we looked for reactions mediated by enzymes that are not cycling at the protein level, harbor cycling phosphorylations, and have rhythmic substrates and/or products. The filtering reduced the network from 6,453 nodes and 48,625 edges to 1,898 nodes and 5,636 edges. The number of remaining edges after filtering should be compared with the number of pairwise comparisons in a naive all-versus-all correlation approach, which was estimated to result in nearly nine million p value calculations. Finally, 52 edges remain that represent pairs of cycling phosphosites and metabolites that are connected through a reaction. For more details on how to perform the filtering, see STAR Methods and Figure 4.

Phosphorylation as driver of dynamic enzymatic reactions

The multi-omics network analysis with Metis resulted in a metabolic sub-network with cycling metabolites of reactions mediated by enzymes with rhythmic phosphorylation changes (Figure 5; Table S1). Together, those rhythmic reactions covered several metabolic processes that involve important metabolites such as NADH, AMP, coenzyme A (CoA), and amino acids. Overall, we find that phosphorylation works as a regulatory switch for enzymes in key metabolic reactions in mouse liver. While phosphorylation can serve as an activating or repressive modulator, most of the time, the functional relevance of a phosphorylated residue identified in a large-scale phosphoproteomics study is unknown (Needham et al., 2019). This is the case for many of the rhythmic phosphorylations in metabolic enzymes of our network as well. We thus seek to infer the functional role of the phosphorylation in the activity of some of these enzymes based on the quantitative correlation with the substrate and/or product of its reaction. An evident example of this is the case of carbamoyl-phosphate synthetase 2 (CAD), which is confirmed by our analysis, where we reproduce the fact that the enzyme is allosterically regulated by phosphorylation at S1859 (Figure S2) (Robitaille et al., 2013).

Moreover, S26 phosphorylation of acyl-CoA oxidase 1 (ACOX1), the enzyme catalyzing the first step of peroxisomal very-long-chain fatty-acid oxidation, cycles with a peak in the resting phase concomitant with the nadir of flavin adenine dinucleotide (FAD), the cofactor of this reaction (Figure 5A, left panel). Low co-factor levels could indicate high enzymatic activity plausibly driven by phosphorylation, leading to a temporal regulation of peroxisomal fatty-acid oxidation with a peak in the inactive phase as reported for mitochondria fatty-acid oxidation (Neufeld-Cohen et al., 2016). FAD is also a co-factor in the first step

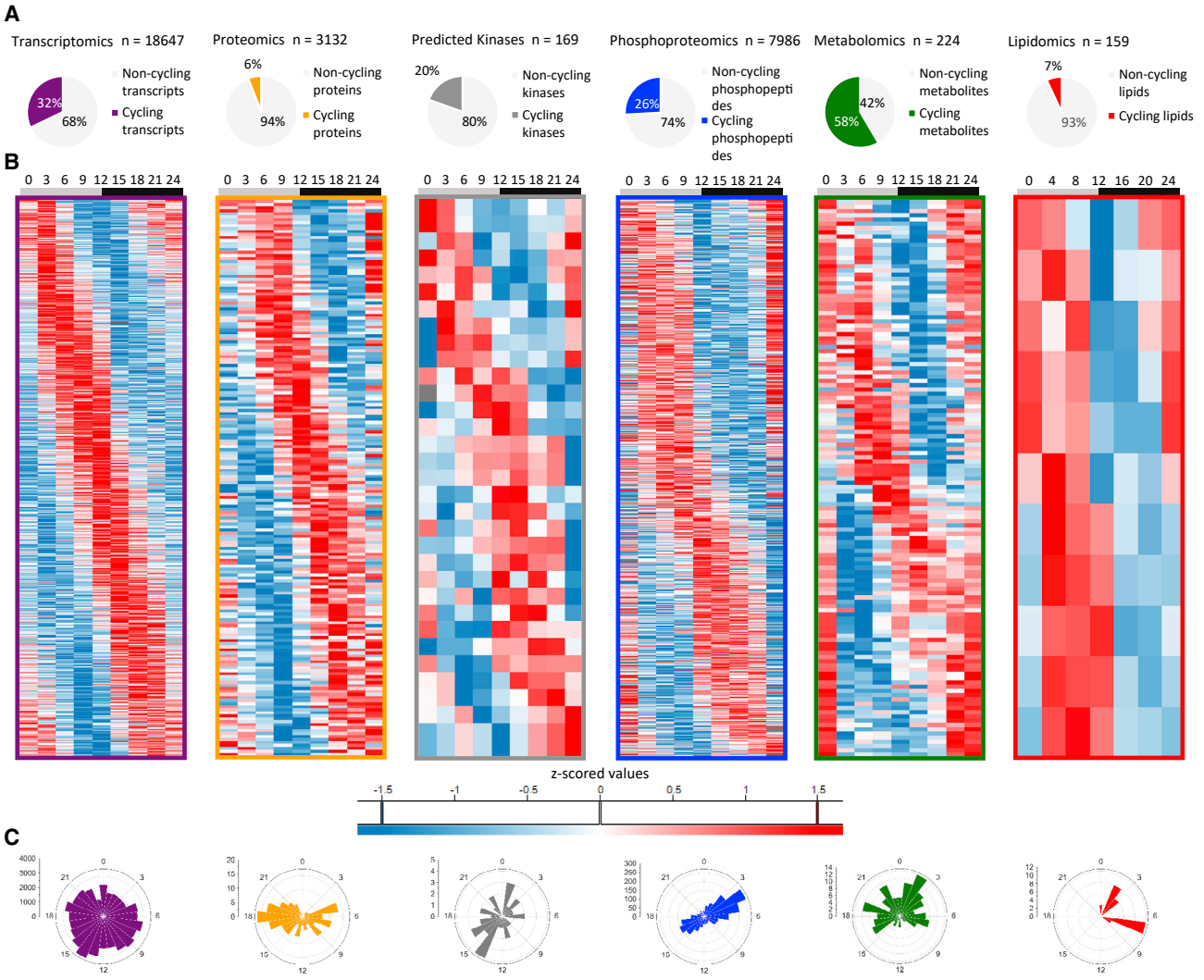


Figure 3. Results of cycling analysis for the individual omics datasets

(A) Pie charts show the percentage of cycling transcripts, proteins, predicted kinases, phosphosites, metabolites, and lipids with respect to the total dataset.

(B) Heatmaps of cycling molecules are shown for each omics type. Biomolecules are sorted vertically according to their circadian phase, while the time points were mapped horizontally and averaged to a 24 h interval in case the time series were longer.

(C) Rose plots indicating peaking positions of all cycling molecules in a circular histogram.

of proline catabolism mediated by proline dehydrogenase (PRODH). In addition to the co-factor FAD, proline as a substrate of this reaction is also rhythmic with a nadir during the inactive phase similar to FAD and also to the cycling profile of PRODH phosphorylation of S32. Thus, under nutrient stress during the resting phase when mice are not eating, PRODH activation would mediate proline catabolism to maintain the cellular energy levels (Pandhare et al., 2009). In contrast to what we observed for ACOX1, increased S32 phosphorylation of PRODH would lead to enzymatic inhibition and accumulation of proline during the active phase when nutrient levels are high due to food intake (Figure 5A, middle and right panels).

Another example is the crosstalk between acetyl-CoA synthase (ACSS2) rhythmic phosphorylation and the cycle of its enzymatic co-factor CoA. ACSS2 is rhythmically phosphorylated in S267, while S30, S263, S267, and S263 phosphorylations do not cycle. Acetyl-CoA is produced by ACSS2 using citrate and CoA as substrates; therefore, the fact that the CoA cycles in anti-phase to the ACSS2 S267 phosphorylation suggests that ACSS2 activity is promoted by S267 phosphorylation (Figure 5A). A similar relationship can be inferred for the CoA synthase COASY, for which cycling phosphorylation at S177 and S182 occurred parallel to the rhythmic levels of its enzymatic product CoA (Figure 5A).

Another very interesting cross correlation between rhythmic enzymatic phosphorylation and cycles of substrate and product metabolites is the reaction mediated by the glycine N-methyltransferase (GNMT). GNMT catalyzes the synthesis of N-methylglycine (sarcosine) from glycine using S-adenosylmethionine (SAM) as the methyl donor, producing SAM. The peak

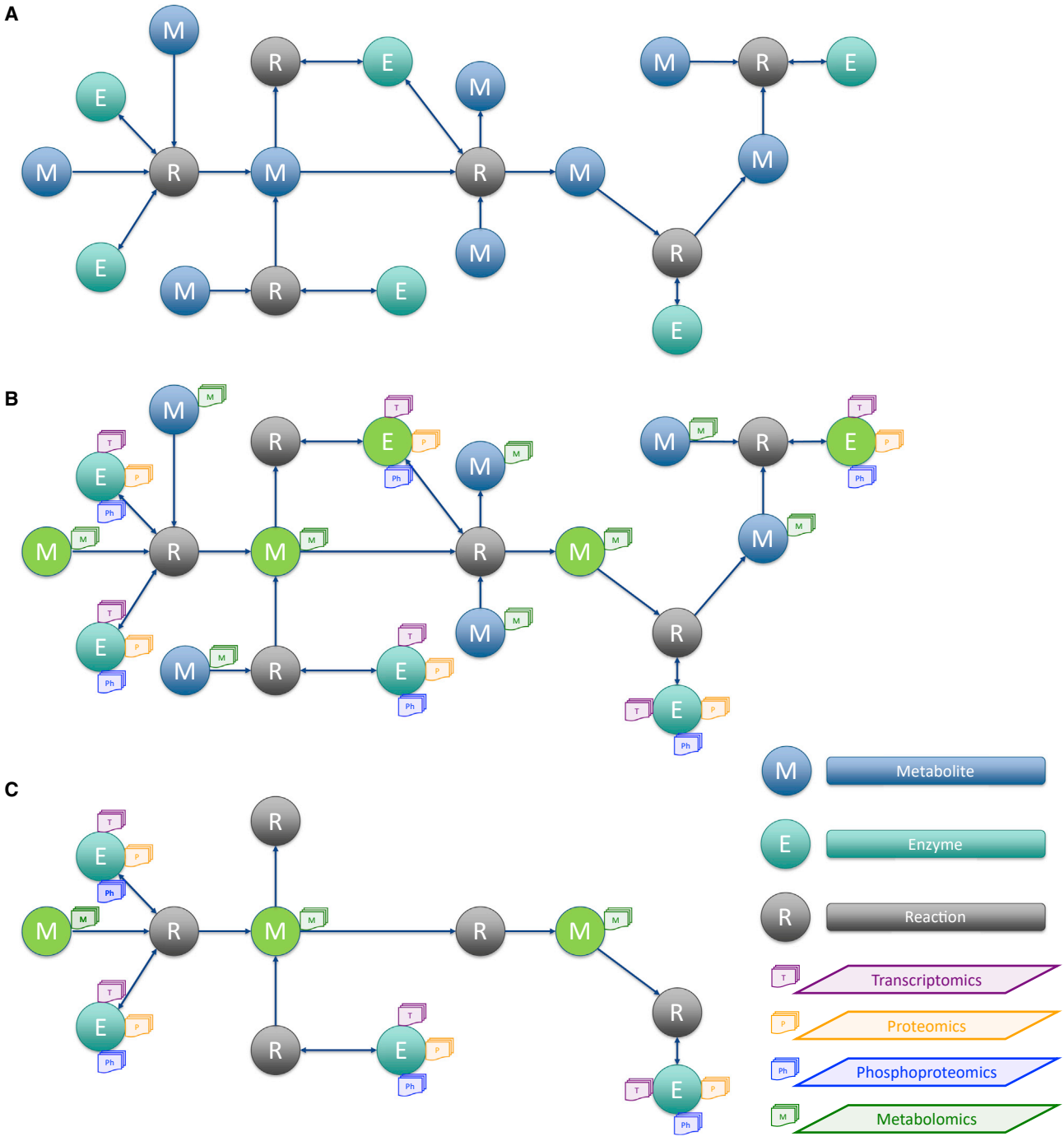
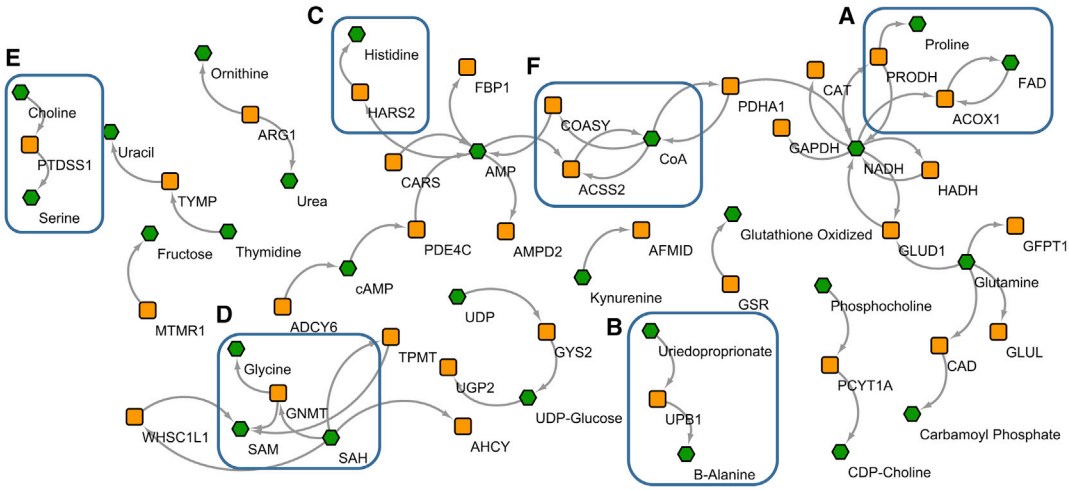


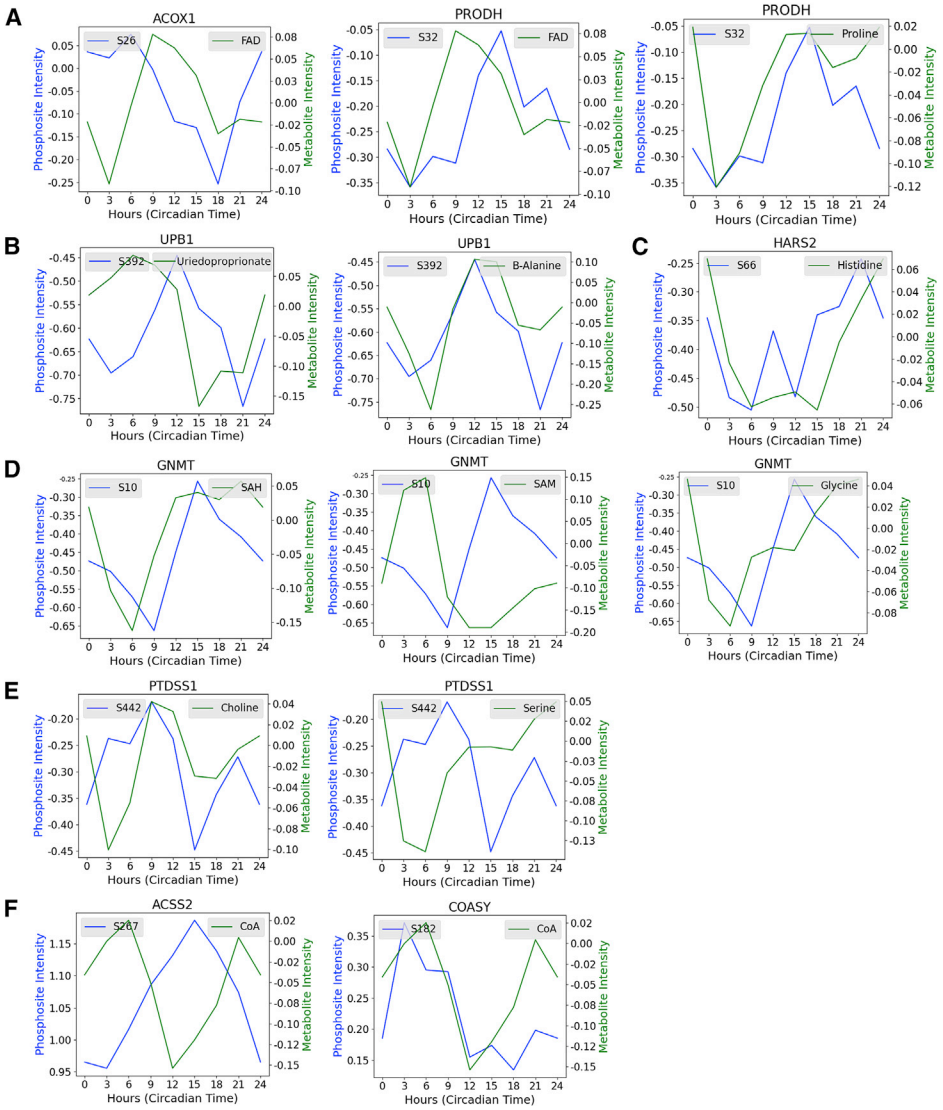
Figure 4. Representation and processing of metabolic networks within Metis

(A) Exemplary network structure that is represented in Metis using the three node types, reaction, metabolite, and enzyme, which are connected according to the participation of biomolecules in reactions.

(B) Nodes are attached to annotations originating from matrices filled with quantitative omics data. The metabolite nodes are affiliated with rows in a matrix object carrying metabolite concentrations over potentially complex experimental designs, which are time series in the case at hand. The enzyme nodes are primarily associated with proteomics data, connecting nodes to the quantitative data on relative enzyme concentrations, here also in the form of time-series data. Also, quantitative post-translational modification data, as, for instance, phosphorylation or mRNA level data, are mapped here to the enzyme nodes. After mapping various omics data to the network, filters can be applied on each node (filtered nodes are shown in green in this example).

(C) Applying the filter results in a network where only the relevant reactions and enzymes are retained.





(legend on next page)

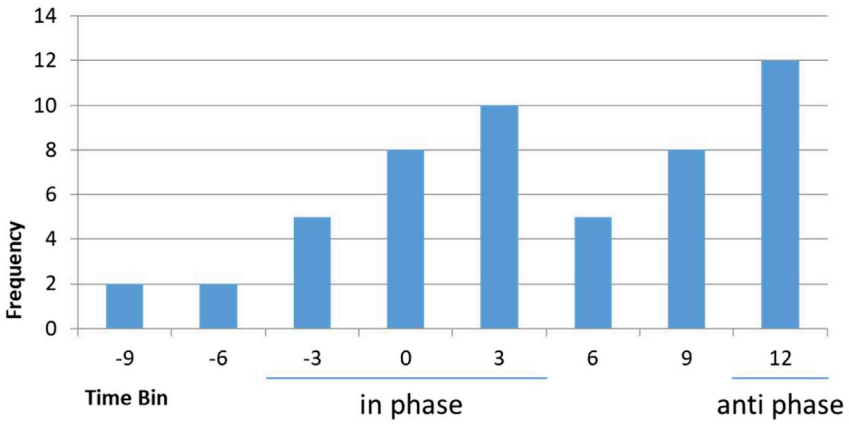


Figure 6. Histogram of the difference between acrophases of metabolites and phosphosites

All pairs of cycling metabolites and phosphorylation sites resulting from network filtering that are hence connected through a reaction were used to create a histogram of phase differences. The differences between acrophases of metabolites and corresponding enzyme phosphorylations are sorted into 3 h bins. Enrichments can be observed around the "in-phase" and "in-anti-phase" regions.

of GNMT phosphorylation at S10 in the middle of the night, likely driven by feeding as previously reported, is concomitant with the maximum levels of SAH and nadir of SAM, the product and substrate of its enzymatic reaction, respectively (Figure 5A). In this manner, the nutrient-dependent regulation of GMNT activity via phosphorylation would impact methionine metabolism and the methyl cycle by controlling the SAM/SAH ratio. Since SAM is the methyl donor for almost all cellular methylation reactions, temporal control of GNMT activity across the day would likely contribute to the daily rhythms of RNA and histone methylation and their crosstalk to the molecular clock (Fustin et al., 2013, 2020; Greco et al., 2020).

Phase relations between metabolite concentrations and enzyme phosphorylation

In the previous section, we looked into specific examples of pairs of cycling metabolites and cycling enzyme phosphorylation that passed our filters targeted at finding enzyme regulation by phosphorylation. In Figure 6, we provide a histogram of phase differences containing all such pairs found by the network analysis. The phase differences are grouped into 3 h bins. The bin at 0 contains those cases for which the enzyme phosphorylation is in phase with the metabolite levels (phase difference between 1.5 and +1.5 h), while the bin at 12 contains those cases for which metabolite and phosphorylation levels are in anti-phase (phase difference between 10.5 and 13.5 h). The highest bin is the one in exact anti-phase at 12. In this bin, the metabolite concentration is high when the fraction of enzymes that are phosphorylated at the respective site is low and vice versa. One needs to distinguish cases in which the metabolite is counted as a product or as a substrate of the reaction. In the case where the metabolite is a product, these can be interpreted as potential cases of enzyme-activity repression by phosphorylation. In the case where the metabolite is a substrate, the interpretation is enzyme activation, since the more a substrate is consumed, the greater the enzyme function driven by higher phosphorylation levels. Accuracy of the data, simplicity of the fit model, and the binning of phases all give some leeway to the phase relationship, which can accommodate 1 to 2 h time lags due to accumulation or consumption times of metabolite concentrations. It should be emphasized that our example data contain metabolite concentrations, not their fluxes. Time lags resulting from this between phosphorylation changes and metabolite concentrations are expected, since the abundance of a metabolite may increase because the rate of its production has increased or because the rate of its consumption has decreased. A more direct correlation would be expected between phosphorylation and flux changes. However, apparently, in our example, these differences were not large enough to completely dilute the phase relationship between metabolites and phosphorylation. When large-scale metabolic flux data are available, these could be added as edge annotations into a Metis model, and the reaction filtering could be based on these. Another caveat is that the association of phases is not the proof of a causal relationship but only points at potentially causal links between biomolecules that ultimately need to be confirmed by other means.

Another interesting region of the histogram is around time lag 0 (bins −3, 0, and 3 h), which also has an increased number of cases compared with the average. Here, phosphorylation levels are close in phase with the metabolite concentration, leading to the opposite interpretation as in the 12 h bin. Here, the cases with substrates are interpreted as enzyme suppression by phosphorylation, while cases with products are interpreted as enzyme activation. Time lags due to accumulation/consumption effects seem to be larger here, as is manifested by the larger spread of phase differences onto the −3 and 3 h bins.

Figure 5. Network of enzymatic reactions with oscillating enzyme phosphorylation, substrates, and products

Orange squares depict enzymes, and green hexagons depict metabolites. For the highlighted regions, common profile plots of normalized phosphosite and metabolite intensities are shown.

(A) ACOX1 and PRODH.

(B) UBP1.

(C) HARS2.

(D) GNMT.

(E) PTDSS1.

(F) ACSS2, PLA2G4A, and COASY.

In summary, the distribution of phase differences between metabolites and corresponding enzyme phosphorylation is non-uniform and indicative of enzyme-activity regulation due to metabolite and phosphorylation crosstalk. Several of these enzyme-metabolite pairs could be rationalized in the previous subsection.

DISCUSSION

We introduced Metis, a Perseus plugin for the joint analysis of multiple omics datasets through metabolic networks. By restricting comparisons between molecular concentration data from two omics levels to the subset defined by co-occurrence in reactions of the metabolic network, we arrive at a manageable problem with a strongly reduced background of false-positive findings. On circadian multi-omics data for mouse liver, we find enzyme phosphorylation-metabolite pairs co-occurring in reactions, which show phase relationships indicative of activation and repression. The circadian multi-omics analysis using Metis highlights phosphorylation as a major regulatory switch of enzymatic activity regulating daily metabolic reactions in mouse liver as already shown for receptor downstream signaling pathways in this same organ (Robles et al., 2017). A potential drawback of the filtering on reactions is the lack of discovery of interactions outside of the ones known in the metabolic network.

We see data quality in terms of completeness of quantification over the whole time series, as well as quantification accuracy as limitations to the data analysis, in particular for the metabolome and phosphoproteome data. The noisiness led us to perform, as a first step, circadian analysis separately in each of the omics levels and do the network-based analysis with the resulting fit parameters. With advances in data quality, it will be possible, as well as of interest, to perform the network analysis with correlations across omics dimensions on the raw time profiles. We see Perseus and Metis as a very suitable framework for this purpose. The different omics levels used in this publication were measured in different labs. While it is remarkable that in spite of this source of variation, the joint analysis with Metis produces meaningful results, we believe that it would be even better to perform all omics measurements in the same lab, with the same time points and preferably derived from the same samples. Also, current technological developments, as, for instance, the application of data-independent acquisition methods to phosphoproteomics and metabolomics, will likely enable other statistical-analysis methods that directly use correlation co-efficients between different omics datasets.

The analysis of circadian multi-omics datasets using Metis highlighted a predominant role for phosphorylation regulating the activity of metabolic enzymes and, consequently, metabolism. Metabolic states crosstalk to the molecular clock to thus ensure proper circadian responses to metabolic changes (Brown, 2016). One mechanism of crosstalk could involve metabolic enzymes that directly modulate circadian transcriptional control by physically interacting with chromatin remodeling systems, reprograming gene expression in response to the metabolic state (Boon et al., 2020; Li et al., 2018). This reprogramming would be largely based on a post-translational mechanism leading to modifications of histones and non-histone proteins to ultimately control their activity. Our analysis of metabolite and enzymatic activity supports this notion and even exposes complementary metabolic reactions that can impact transcription. For example, while an ACSS2 phospho-dependent peak of activity in the middle of the night would promote histone acetylation (Mews et al., 2017) and transcriptional activity by generating Acetyl-CoA, food-driven phosphorylation and activation of GNMT in the night would inhibit histone methylation, and thus transcriptional repression, by reducing SAM levels (Fustin et al., 2013, 2020; Greco et al., 2020). Thus, rhythms of metabolic enzymatic activity and corresponding metabolite levels would specifically impact the circadian molecular machinery at the chromatin to ultimately entrain the molecular clock to metabolic and nutrient states.

Limitations of the study

While the developed software is generically applicable to the data types presented, limitations in our particular study arose due to the heterogeneity of the data used. Likely more findings would be obtained if all omics data were acquired from the same samples with same sampling of the time axis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Transcriptomics data
  - Proteomics data
  - Phosphoproteomics data
  - Metabolomics data
  - Lipidomics data
  - Kinase activity prediction
  - Periodicity analysis
  - Whole genome metabolic networks
  - Network mapping and filtering
  - Network export to third-party software
  - Mendeley data
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100198.

ACKNOWLEDGMENTS

This project was partially funded by the German Ministry for Science and Education (BMBF) funding action MSCoreSys, reference number FKZ 031L0214D. M.S.R. is funded by the DFG (project 329628492 – SFB 1321 and 428041612) and by LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative. We thank Dr. Jan Daniel Rudolph for help with PHOTON.

AUTHOR CONTRIBUTIONS

H.H. and J.C. designed the tools used within the Perseus software, and H.H. and D.F. implemented them. H.H., M.S.R., and J.C. performed the analyses. M.S.R. and J.C. directed the research. H.H., M.S.R., and J.C. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 25, 2021
Revised: January 14, 2022
Accepted: March 28, 2022
Published: April 14, 2022

REFERENCES

Adamovich, Y., Rousso-Noori, L., Zwighaft, Z., Neufeld-Cohen, A., Golik, M., Kraut-Cohen, J., Wang, M., Han, X., and Asher, G. (2014). Circadian clocks and feeding time regulate the oscillations and levels of hepatic triglycerides. Cell Metab. *19*, 319–330.

Boon, R., Silveira, G.G., and Mostoslavsky, R. (2020). Nuclear metabolism and the regulation of the epigenome. Nat. Metab. *2*, 1190–1203.

Brown, S.A. (2016). Circadian metabolism: from mechanisms to metabolomics and medicine. Trends Endocrinol. Metab. *27*, 415–426.

Brüning, F., Noya, S.B., Bange, T., Koutsouli, S., Rudolph, J.D., Tyagarajan, S.K., Cox, J., Mann, M., Brown, S.A., and Robles, M.S. (2019). Sleep-wake cycles drive daily dynamics of synaptic phosphorylation. Science *366*, eaav3617.

Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. *21*, 630–644.

Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., Mittag, F., Schubert, M., Glont, M., and Golebiewski, M.M. (2013). Path2-Models: large-scale generation of computational models from biochemical pathway maps. BMC Syst. Biol. *7*, 116.

Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., and Burdick, J.T. (2005). Mapping determinants of human gene expression by regional and genome-wide association. Nature *437*, 1365–1369.

Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics. *13*, S12.

Finger, A.M., Dibner, C., and Kramer, A. (2020). Coupled network of the circadian clocks: a driving force of rhythmic physiology. FEBS Lett. *594*, 2734–2769.

Fustin, J.M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M.S., Kakeya, H., Manabe, I., et al. (2013). XRNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell *155*, 793–806.

Fustin, J.M., Ye, S., Rakers, C., Kaneko, K., Fukumoto, K., Yamano, M., Versteven, M., Grünewald, E., Cargill, S.J., Tamai, T.K., et al. (2020). Methylation deficiency disrupts biological rhythms from bacteria to humans. Commun. Biol. *3*, 1–14.

Greco, C.M., Cervantes, M., Fustin, J.M., Ito, K., Ceglia, N., Samad, M., Shi, J., Koronowski, K.B., Forne, I., Ranjit, S., et al. (2020). S-adenosyl-l-homocysteine hydrolase links methionine metabolism to the circadian clock and chromatin remodeling. Sci. Adv. *6*, eabc5629.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512–D520.

Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., Panda, S., and Hogenesch, J.B. (2009). Harmonics of circadian gene transcription in mammals. PLoS Genet. *5*, e1000442.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering *9*, 90–95.

Krishnaiah, S.Y., Wu, G., Altman, B.J., Growe, J., Rhoades, S.D., Coldren, F., Venkataraman, A., Olarerin-George, A.O., Francey, L.J., Mukherjee, S., et al. (2017). Clock regulation of metabolites reveals coupling between transcription and metabolism. Cell Metab. *25*, 1206–1974.e4.

Li, X., Egervari, G., Wang, Y., Berger, S.L., and Lu, Z. (2018). Regulation of chromatin and gene expression by metabolic enzymes and metabolites. Nat. Rev. Mol. Cell Biol. *19*, 563–578.

Mews, P., Donahue, G., Drake, A.M., Luczak, V., Abel, T., and Berger, S.L. (2017). Acetyl-CoA synthetase regulates histone acetylation and hippocampal memory. Nature *546*, 381–386.

Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. (2004). Genetic analysis of genome-wide variation in human gene expression. Nature *430*, 743–747.

Needham, E.J., Parker, B.L., Burykin, T., James, D.E., and Humphrey, S.J. (2019). Illuminating the dark phosphoproteome. Sci. Signal. *12*, eaau8645.

Neufeld-Cohen, A., Robles, M.S., Aviram, R., Manella, G., Adamovich, Y., Ladeuix, B., Nir, D., Rousso-Noori, L., Kuperman, Y., Golik, M., et al. (2016). Circadian control of oscillations in mitochondrial rate-limiting enzymes and nutrient utilization by PERIOD proteins. Proc. Natl. Acad. Sci. U. S. A. *113*, E1673–E1682.

Panda, S., Antoch, M.P., Miller, B.H., Su, A.I., Schook, A.B., Straume, M., Schultz, P.G., Kay, S.A., Takahashi, J.S., and Hogenesch, J.B. (2002). Coordinated transcription of key pathways in the mouse by the circadian clock. Cell *109*, 307–320.

Pandhare, J., Donald, S.P., Cooper, S.K., and Phang, J.M. (2009). Regulation and function of proline oxidase under nutrient stress. J. Cell. Biochem. *107*, 759–768.

Robitaille, A.M., Christen, S., Shimobayashi, M., Cornu, M., Fava, L.L., Moes, S., Prescianotto-Baschong, C., Sauer, U., Jenoe, P., and Hall, M.N. (2013). Quantitative phosphoproteomics reveal mTORC1 activates de novo pyrimidine synthesis. Science *339*, 1320–1323.

Robles, M.S., Cox, J., and Mann, M. (2014). In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. PLoS Genet. *10*, e1004047.

Robles, M.S., Humphrey, S.J., and Mann, M. (2017). Phosphorylation is a central mechanism for circadian control of metabolism and physiology. Cell Metab. *25*, 118–127.

Rudolph, J.D., and Cox, J. (2019). A network module for the perseus software for computational proteomics facilitates proteome interaction graph analysis. J. Proteome Res. *18*, 2052–2064.

Rudolph, J.D., de Graauw, M., van de Water, B., Geiger, T., and Sharan, R. (2016). Elucidation of Signaling pathways from large-scale phosphoproteomic data using protein interaction networks. Cell Syst. *3*, 585–593.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498–2504.

Storch, K.F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F.C., Wong, W.H., and Weitz, C.J. (2002). Extensive and divergent circadian gene expression in liver and heart. Nature *417*, 78–83.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods. *13*, 731–740.

Ueda, H.R., Chen, W., Adachi, A., Wakamatsu, H., Hayashi, S., Takasugi, T., Nagano, M., Nakahama, K.I., Suzuki, Y., Sugano, S., et al. (2002). A transcription factor response element for gene expression during circadian night. Nature *418*, 534–539.

van der Walt, Colbert, S., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science & Engineering *13*, 22–30.

Yu, S.H., Ferretti, D., Schessner, J.P., Rudolph, J.D., Borner, G.H.H., and Cox, J. (2020). Expanding the perseus software for omics data analysis with custom plugins. Curr. Protoc. Bioinforma. *71*, e105.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Mouse liver circadian transcriptomics data	(Hughes et al., 2009)	https://doi.org/10.1371/journal.pgen.1000442
Mouse liver circadian proteomics data	(Robles et al., 2014)	https://doi.org/10.1371/journal.pgen.1004047
Mouse liver circadian phosphoproteomics data	(Robles et al., 2017)	https://doi.org/10.1016/j.cmet.2016.10.004
Mouse liver circadian metabolomics data	(Krishnaiah et al., 2017)	https://doi.org/10.1016/j.cmet.2017.03.019
Mouse liver circadian lipidomics data	(Adamovich et al., 2014)	https://doi.org/10.1016/j.cmet.2013.12.016
Whole Genome Metabolism - *Mus musculus*	(Büchel et al., 2013)	https://doi.org/10.1186/1752-0509-7-116
Software and algorithms		
Perseus	(Tyanova et al., 2016) and this paper.	https://doi.org/10.1038/nmeth.3901
Cytoscape	(Shannon et al., 2003)	https://doi.org/10.1101/gr.1239303
Perseus sessions	This paper.	https://doi.org/10.17632/n4nx6x999v.1
Matplotlib	(Hunter, 2007)	https://doi.org/10.1109/MCSE.2007.55
Pandas		http://www.dlr.de/sc/Portaldata/15/Resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf
NumPy	(van der Walt et al., 2011)	https://doi.org/10.1109/MCSE.2011.37

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jürgen Cox, cox@biochem.mpg.de.

Materials availability

This work consists purely of data and data analysis software for the generation of the results and thus no further materials were used.

Data and code availability

- All re-analyzed data have been deposited and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table. Data analysis was performed with Perseus version 1.6.15 containing the Metis plugin, which can be downloaded from https://maxquant.org/perseus/and whose code is freely available on GitHub (https://github.com/JurgenCox/perseus-plugins/tree/master/PluginMetis). The Metis plug-in is distributed under the GNU v3.0 license. All Perseus sessions and further data are available in the Mendeley dataset https://doi.org/10.17632/n4nx6x999v.1.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Transcriptomics data

The mouse liver circadian transcriptomics data contained 18,647 transcripts, quantified every hour (48 time points). The series matrix was downloaded as a.txt file from the Gene Expression Omnibus (GEO) with the ID GSE11923 (Hughes et al., 2009). The header of the downloaded file was removed, only keeping the ''!Sample_title''. This file was then uploaded to Perseus using the ''Generic matrix upload'' function where all columns containing expression values were uploaded as ''Main'' and the ''ID_REF'' column was uploaded as ''Text''. The ''ID_REF'' column was used to annotate the transcripts with UniProt IDs using a Perseus annotation file shipped with the software (which was also uploaded to Perseus using the ''Generic matrix upload'' function) and the ''Matching rows by name'' function of Perseus. Rows which were not annotated with an UniProt ID were removed using the ''Filter rows based on text column'' function. The rows with the same UniProt ID were combined, taking the median using the ''Unique rows'' function. Having checked

the distribution of the data using the "Histogram" function, the data was transformed by log10(x) using the "Trans-form" function (see Perseus session file named "Transcriptomics.sps").

Proteomics data

The mouse liver circadian proteomics data contained 3,132 proteins, quantified every 3 h (16 time points). The data was obtained from the supplementary material of the original article (Robles et al., 2014) as an Excel document. The sheet named "A- Total dataset" was saved as a.txt file and uploaded to Perseus using the "Generic matrix upload" function where all columns containing expression values were uploaded as "Main" and the UniProt IDs of the proteins as "Text". Having checked the distribution of the data using the "Histogram" function, and since the original paper reports the expression values to have been transformed by log2(x), the data was first transformed by $2^{(x)}$ and later by log10(x) using the "Trans-form" function. This is done so that all datasets are treated exactly in the same manner within Perseus (see Perseus session file named "Proteomics.sps"). While the protein quantification data from this source was produced with MaxQuant, Metis and Perseus in general operate independently of the MaxQuant software and can process omics data produced by any other software.

Phosphoproteomics data

The mouse liver phosphoproteomics data contained 7,986 phosphosites quantified every 3 h (16 time points). The data was obtained from the supplementary material of the original article (Robles et al., 2017) as an Excel document. The sheet named "A- Total dataset" was saved as a.txt file and uploaded to Perseus using the "Generic matrix upload" function where all columns containing expression values were uploaded as "Main", the UniProt IDs of the proteins as "Text", the phosphorylated amino acid as "Text" and the position of the phosphorylated amino acid within the protein as "Text". Since the data was already log10(x) transformed, no further processing was carried out (see Perseus session file named "Phosphoproteomics.sps").

Metabolomics data

The mouse liver circadian metabolomics data contained 224 metabolites, quantified every hour (48 time points). The data was obtained from the supplementary material of the original article (Krishnaiah et al., 2017) as an Excel document. The sheet named "Liver_data" was saved as a.txt file. For the measured metabolites we could retrieve 200 ChEBI IDs which were later used to map and annotate the metabolites according to the mouse metabolic network from the BioModels database. The ChEBI IDs were retrieved via a simple script from HMDB website using the supplied HMDB IDs within the original supplementary file provided by Krishnaiah et al. (2017). For metabolites missing HMDB IDs, metabolite names were used for manual retrieval of ChEBI IDs. The resulting.txt file was then uploaded to Perseus using the "Generic matrix upload" function where all columns containing the metabolite quantification values were uploaded as "Main". Having checked the distribution of the data using the "Histogram" function, the data was transformed by log10(x) using the "Trans-form" function (see Perseus session file named "Metabolomics.sps").

Lipidomics data

The lipidomics data contained 159 lipids, quantified every 4 h (6 time points). The data was obtained from the supplementary material of the original article (Adamovich et al., 2014) as an Excel document. The sheet named "A." was modified to have the lipid types as a column instead of row separators and saved as a.txt file and uploaded to Perseus using the "Generic matrix upload" function where all columns containing the lipid quantification values were uploaded as "Main", the mass as "Numeric" and the type and name as "Text". Having checked the distribution of the data using the "Histogram" function, the data was transformed by log10(x) using the "Trans-form" function (see Perseus session file named "Lipidomics.sps").

Kinase activity prediction

Kinase activity prediction was performed using the PHOTON plugin of Perseus. The phosphoproteomics data was annotated based on the UniProt IDs with Gene IDs and ENSP IDs using the Perseus "Add annotation" function. Three different protein-protein interaction networks were used, namely BioGRID, IntAct and STRING. These were downloaded from the respective web sources and are available as supplementary data within this paper. After the analysis described below for each interaction network, the periodicity analysis is performed as explained in the Periodicity Analysis section and the resulting matrices are merged, annotated using the "Add annotation" function with "Keywords" which were later filters for "Kinase" using the "Filter rows based on categorical column" to keep only the kinases from the PHOTON prediction (see Perseus session file named "Kinase Activity Prediction.sps").

The BioGRID network.txt file and uploaded to Perseus using the "Generic matrix upload" function with the confidence column as "Numeric" and the source and target columns as "Text". The resulting matrix was converted to the Perseus "Network collection" data type using the "From matrix" function of Perseus choosing the correct "Source" and "Target" columns. Then the node degrees were calculated using the "Node degrees" function of Perseus and filtered using the "Filter nodes by numerical column" function for nodes with less than 600° in order to discard nodes that are connected to too many other nodes which would cause significant noise within the PHOTON analysis. The remaining nodes within the network were then annotated with the phosphoproteomics quantitative time series data (all the "Main" columns) using the "Annotate nodes" function based on the "Node" column of the network and the "GeneID" column of the phosphoproteomics data, selecting "Keep separate" for the "Combine copied main values" option. PHOTON analysis was done using the PHOTON plugin of Perseus where all the columns contacting the quantitative data are selected

and the ''Side'' is selected as ''twosided'' and a python.exe path is given as ''Executable''. As output, PHOTON provides two network collections and a matrix. The matrix was further processed using the ''To base identifiers'' function based on the ''Node'' column containing the Gene IDs to retrieve the UniProt IDs for the PHOTON prediction. The columns were also renamed using the ''Rename columns'' function and sorted using the ''Sort columns'' function. The resulting matrix was saved as a.txt file using the ''Generic matrix export'' function.

The IntAct network.txt file and uploaded to Perseus using the ''Generic matrix upload'' function with the confidence column as ''Numeric'' and the source and target columns as ''Text''. The resulting matrix was filtered using the ''Filter nodes by numerical column'' function for the protein-protein interactions with confidence values greater than 0.5. The resulting converted to the Perseus ''Network collection'' data type using the ''From matrix'' function of Perseus choosing the correct ''Source'' and ''Target'' columns. The nodes within the network were then annotated with the phosphoproteomics quantitative time series data (all the ''Main'' columns) using the ''Annotate nodes'' function based on the ''Node'' column of the network and the column containing the UniProt IDs of the phosphoproteomics data, selecting ''Keep separate'' for the ''Combine copied main values'' option. PHOTON analysis was done using the PHOTON plugin of Perseus where all the columns contacting the quantitative data are selected and the ''Side'' is selected as ''twosided'' and a python.exe path is given as ''Executable''. As output, PHOTON provides two network collections and a matrix. The matrix was further processed to rename the columns using the ''Rename columns'' function and sort them using the ''Sort columns'' function. The resulting matrix was saved as a.txt file using the ''Generic matrix export'' function.

The STRING network.txt file was uploaded to Perseus using the ''Raw upload'' function with the ''Split into columns'' selected along with the ''Separator'' as ''Tab''. The resulting matrix's score column type was converted from string to numerical using the ''Change column type'' function. Later the score column was used to filter for interactions with a score higher than 900 using the ''Filter nodes by numerical column'' function. The ''mode'' column was also used to filter for protein-protein interactions which were categorized as ''binding'' using the ''Filter rows based on text column'' function with the search string set as ''binding'', without matching for case but matching for the whole word and the ''Mode'' selected as ''Keep matching rows'' and the filter mode as ''Reduce matrix''. The resulting matrix was further processed using the ''Process text column'' function to remove ''10,090.'' from the beginning of the ENS IDs within the STRING network ''item_id_a'' and ''item_id_b'' columns with the regular expression ''10,090\.(.*)'' and no replacement string. The ''Rename columns'' function was used to rename the columns within the matrix containing the ENS IDs to ''protein1'' and ''protein2'' for the interacting proteins. The ''Trans-form'' function was then used on the confidence column with the transformation formula as ''x/1000''. Later, using the ''Reorder/remove columns'' function, only the columns ''Confidence'', ''protein1'' and ''protein2'' were kept for creating the Perseus network collection using the ''From matrix'' function to be used for PHOTON analysis. Prior to the PHOTON analysis, the node degrees were calculated using the ''Node degrees'' function which was used to filter the network for nodes with degrees less than 1000 in order to discard nodes that are connected to too many other nodes which would cause significant noise within the PHOTON analysis. The nodes within the network were then annotated with the phosphoproteomics quantitative time series data (all the ''Main'' columns) using the ''Annotate nodes'' function based on the ''Node'' column of the network and the column containing the ENSP IDs of the phosphoproteomics data, selecting ''Keep separate'' for the ''Combine copied main values'' option. PHOTON analysis was done using the PHOTON plugin of Perseus where all the columns contacting the quantitative data are selected and the ''Side'' is selected as ''greater'' and a python.exe path is given as ''Executable''. As output, PHOTON provides two network collections and a matrix. The matrix was further processed to rename the columns using the ''Rename columns'' function and sort them using the ''Sort columns'' function. The resulting matrix was saved as a.txt file using the ''Generic matrix export'' function.

Periodicity analysis

All the data acquired from previous publications and the kinase activity prediction were analyzed using Perseus's time-series analysis toolkit. Perseus performs this analysis in a permutation-based FDR-controlled manner and calculates the amplitude of the change and the peaking time for each case by fitting the data to a cosine function (Tyanova et al., 2016). The results were then filtered to define ''cycling'' and ''non-cycling'' entries according to the original q-values recommended by each of the publications that the data originated from and q-value less than 0.1 for the kinase activity prediction data using the ''Filter rows based on numerical/main column'' function. The data was annotated using the ''Categorical annotation rows'' and ''Numerical annotation rows'' functions for all the measurements within the 48 h. The last five time points of the transcriptomics data were removed due to systematic abnormalities observed in the data using the ''Reorder/remove columns'' function. Using the numerical annotation, the circadian analysis was done using the ''Periodicity analysis'' function with the period set to 23.6, FDR set to 1 and number of randomizations set to 1000. The heat maps were made by collapsing and averaging the measurements to 3 h intervals using ''Average groups'' function and the zero and 24 time point were calculated using the same data points. Prior to visualizing the heatmap the ''Z-score'' function was used for normalization and the ''Hierarchical clustering'' function was used without the row and column trees (see Perseus session files named ''Lipidomics.sps'', ''Metabolomics.sps'', ''Phosphoproteomics.sps'', ''Proteomics.sps'', ''Transcriptomics.sps'' and ''Kinase Activity Prediction Cycling Analysis.sps'').

Whole genome metabolic networks

The metabolic network used in this study and available at http://annotations.perseus-framework.org/within the ''MetabolicNetworks'' folder for 11 most popular model organisms are based on data downloaded from the BioModels database (Path2Models)

(Büchel et al., 2013). These metabolic networks are parsed upon retrieval from the BioModels database as an.xml file and reduced to two text files containing the edges and the node annotations of the network. The.txt file with the edges of the network contains two columns, ''Source'' and ''Target'', while the.txt file with the node annotations contains two columns, ''Node'' and ''Type''. These files can simply be uploaded to Perseus as explained in the following section and used.

Network mapping and filtering

The network.txt files for mouse were uploaded to Perseus using the ''Generic matrix upload'' function. For the edge table both the ''Source'' and ''Target'' columns are selected as ''Text'' and for the node annotation table the ''Node'' column is selected as ''Text'' and the ''Type'' column as ''Categorical''. Using the ''From matrix'' function, the matrix containing the edges of the network was converted to a Perseus network collection. Later using the matrix containing the node annotations, annotations were added to the network using the ''Annotate nodes'' function. Note that these annotations can also be extended using the Perseus annotation files shipped with the software and also available at http://annotations.perseus-framework.org/within the ''PerseusAnnotation'' folder. For the purpose of this study we annotate the mouse metabolic network with q-values from each of the omics dataset for which the periodicity analysis was performed using Perseus as explained in the previous sections. The matrix from each periodicity analysis performed on the various omics datasets was exported in.txt format and imported into the Perseus session containing the metabolic network collection using the ''Generic matrix upload'' function. Since these.txt files are generated using Perseus, upon upload to another Perseus session, Perseus assigns the correct data types to each column in the file automatically. In order to map the q-values from the metabolomics data to the network the ChEBI IDs were used but since there were formatting differences between the network IDs and the data, the ''Process text column'' function was used to add the string ''CHEBI:'' to the beginning of the IDs within the metabolomics data with the regular expression ''^([^]+)'' and the replacement string ''CHEBI:$&'' on the ''ChEBI'' column of the matrix. The column containg the q-values was also renamed using the ''Rename columns'' function to ''Metabolomics q-value'' prior to using the ''Annotate nodes'' function for mapping the data. The same strategy is used to map the q-values from the phosphoproteomics, proteomics and transcriptomics data. After annotating the network with the q-values, the network was filtered using the ''Filter for metabolic reactions'' function with the ''Number of columns'' set to four, ''x'' set to the metabolomics q-values, ''y'' set to the phosphoproteomics q-values, ''z'' set to the proteomics q-values and ''a'' can be set to the transcriptomics q-values. Subsequently, the four relations are set as ''x < 0.05'', ''y < 0.1'' and ''z>=0.33'' and no restriction on a. The ''Combine through'' option is set to ''union'', since each node type is annotated with a separate column containing the q-values of the cycling analysis for each of the relevant datasets. This results in reactions where either of the three cases is true (cycling metabolite/s, cycling phosphosite/s and non-cycling protein/s). The filtering reduced the network from 6,453 nodes and 48,625 edges to 1,898 nodes and 5,636 edges. For further analysis of the remaining reactions, the ''Metabolic reactions to matrix'' function was used to collapse each reaction with the filtered nodes to the Perseus matrix format where each row represents a reaction. For this purpose, the column containing the node types were selected, also, the reaction, modifier (enzyme/protein), substrate and products were selected. The resulting matrix is then further processed to filter for reactions that have both phosphosite information for their modifiers and metabolite information for their reactants to reach phosphosite and metabolite pairs of interest (see the Perseus session file named ''Analysis-Time-Series.sps).

Network export to third-party software

Any metabolic network analyzed within Perseus can be exported in the.sif (simple interaction file) format using the ''SIF export for metabolic reactions'' function. For this purpose, both the Perseus metabolic network matrix and network types can be used. For exporting networks from the matrix format, the columns containing the reaction, modifier (enzyme/protein), substrate and products need to be selected along with the path to a Python installation. Exporting the metabolic network from Perseus network format, simply use the sif export function in the network tab (Figure S3). The resulting matrix can then be exported and used within third-party software, e.g. Cytoscape.

Mendeley data

Hamzeiy, Hamid; Ferretti, Daniela; Robles, Maria; Cox, Juergen (2021), ''Perseus Metis Data'', Mendeley Data, V1, https://doi.org/10.17632/n4nx6x999v.1 https://doi.org/10.17632/n4nx6x999v.1

Metabolic Network Analysis Folder: Perseus session files and data used for Metis analysis for both time series data and static data, along with the final results presented.

Metabolic Networks Folder: Metabolic networks for 11 most common organisms.

Periodicity Analysis Folder: Perseus session files and data used to perform the periodicity analysis on the various datasets. PHOTON Analysis Folder: Protein-protein interaction networks and Perseus session files used to perform PHOTON analysis for kinases activity prediction.

QUANTIFICATION AND STATISTICAL ANALYSIS

No applicable statistical tests have been performed.