Journal of proteome research

pubs.acs.org/jpr

Technical Note

PHONEMeS: Efficient Modeling of Signaling Networks Derived from Large-Scale Mass Spectrometry Data

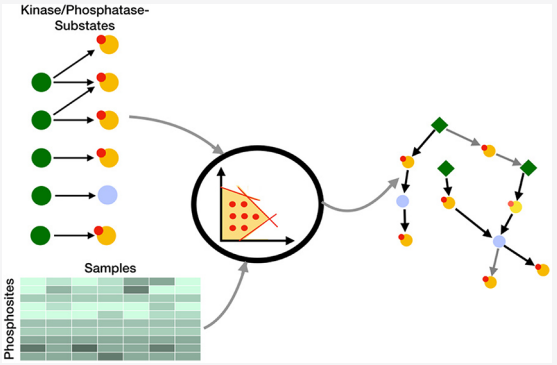
Enio Gjerga, Aurelien Dugourd, Luis Tobalina, Abel Sousa, and Julio Saez-Rodriguez*

Cite This: *J. Proteome Res.* 2021, 20, 2138–2144

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Post-translational modifications of proteins play an important role in the regulation of cellular processes. The mass spectrometry analysis of proteome modifications offers huge potential for the study of how protein inhibitors affect the phosphosignaling mechanisms inside the cells. We have recently proposed PHONEMeS, a method that uses high-content shotgun phosphoproteomic data to build logical network models of signal perturbation flow. However, in its original implementation, PHONEMeS was computationally demanding and was only used to model signaling in a perturbation context. We have reformulated PHONEMeS as an Integer Linear Program (ILP) that is orders of magnitude more efficient than the original one. We have also expanded the scenarios that can be analyzed. PHONEMeS can model data upon perturbation on not only a known target but also deregulated pathways upstream and downstream of any set of deregulated kinases. Finally, PHONEMeS can now analyze data sets with multiple time points, which helps us to obtain better insight into the dynamics of the propagation of signals. We illustrate the value of the new approach on various data sets of medical relevance, where we shed light on signaling mechanisms and drug modes of action.



KEYWORDS: *phosphoproteomics, cell signaling, modelling, integer linear programming*

INTRODUCTION

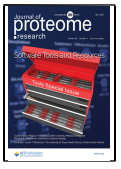
The recent developments in mass spectrometry (MS) technologies have enormously advanced the study of the proteome and its role in the functioning of cells. The high degree of coverage of the proteome and the ever-improving accuracy due to technical advancements have made shotgun MS one of the most frequently used proteomic discovery methods.[1] Whereas untargeted MS is limited by low quantitative accuracy and missing data, the superior coverage of this approach offers a huge potential for the identification of additional signaling elements responsible for processes that might not be fully understood. However, the interpretation of untargeted large-scale data sets remains a major challenge in the field.

For the analysis of MS data, users can rely on statistical methods such as differential expression and enrichment analysis to extract interpretable information (such as deregulated pathways) from these large data sets. Pathway enrichment analysis can find correlations between regulated signaling components but ignores interactions between proteins. To gain mechanistic insights, we can integrate the data with known biological processes represented as networks of interacting proteins. This allows us to apply a wide range of systems biology methods to simulate and analyze the deregulations of cell signaling. For this, several tools such as PHOTON,[2] Temporal Pathway Synthesizer,[3] and PHONEMeS (PHOsphorylation NEtworks for Mass Spectrometry)[4] have been developed to reconstruct signaling pathways from large MS phosphoproteomic data.

In particular, PHONEMeS was developed as a method to train and build large-scale Boolean logic models of signaling. PHONEMeS combines identified phosphopeptide abundance measurements (e.g., from untargeted shotgun phosphoproteomic experiments) with a large Prior Knowledge Network (PKN). The PKN is primarily built from known kinase/phosphatase-to-substrate (K/P−S) relations. PHONEMeS identifies signaling pathways in the PKN connecting a priori known perturbation targets to differentially regulated phosphopeptide measurements within the PKN. In brief, the PKN is mapped into an interaction graph that consists of nodes (representing the enzymes and phospho sites) and interactions (the known K/P−S relations). Then, through a stochastic optimization approach, PHONEMeS tries to identify a subset of functional kinase-to-substrate interactions based on the evidence from the data. However, the method was computa-

Special Issue: Software Tools and Resources 2021

Received: November 30, 2020
Published: March 8, 2021



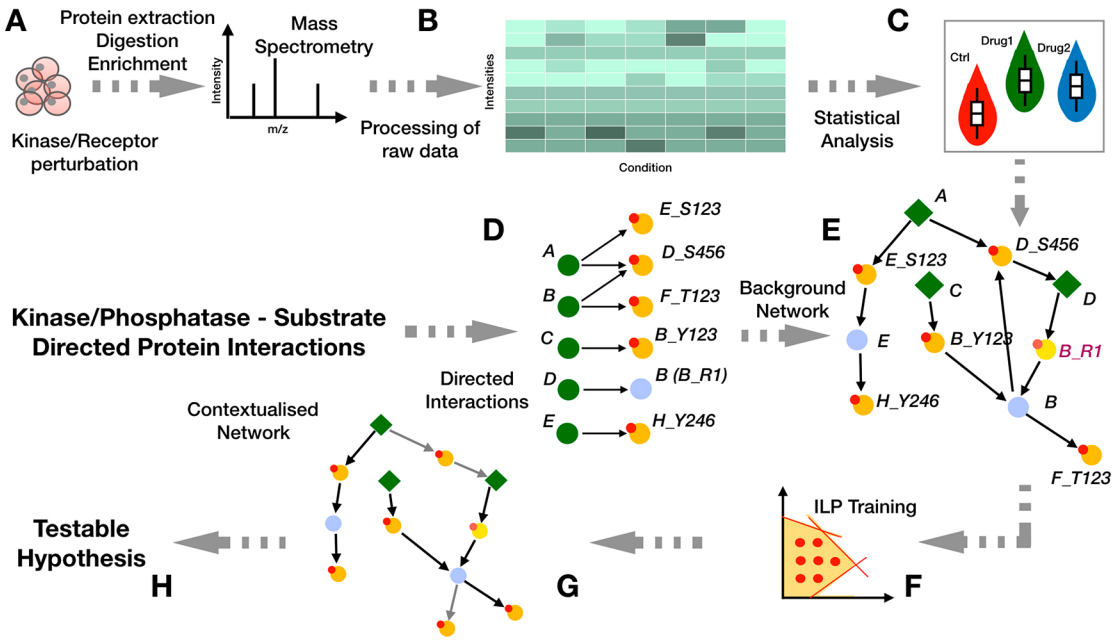


Figure 1. PHONEMeS pipeline. (A) Protein extraction from treated cells. (B) Phosphoproteomic mass spectrometry analysis. (C) Statistical analysis and assignment of scores to each measurement based on statistical evidence. (D) Assembly of a PKN from kinase-to-substrate and directed protein relations. Green nodes represent proteins that can serve as enzymes. Nodes with small red circles on the top left side represent proteins that can be phosphorylated. Light-blue nodes represent intermediate nodes that can be affected by another protein through other mechanisms besides phosphorylation/dephosphorylation. (E) From a bipartite network, we obtain the prior knowledge network that is used for the training. This is done by connecting our perturbation targets (diamond green nodes: A, C, and D in this case) with measured phosphosites (yellow nodes with small red circles at the top left) at the protein level and then translating the directed interactions between proteins into their protein-site equivalents. This is achieved by the addition of integrator nodes that propagate the information from a kinase site to the kinase itself (i.e., nodes E, D, and B). This is the network that is used for training. For interactions in which we have no information about the site that is being targeted (i.e., DB), we model it by designating the target node with an auxiliary site that has a recognizable pattern "_R1" (i.e., node B is depicted on the bipartite network as B_R1). (F) Training of the PKN to data through an ILP formulation. (G) Reporting the contextualized network solution. (H) Obtaining insights from PHONEMeS network results.

tionally demanding and was limited only to handling data sets in a chemical perturbation context (that is, data sets with targeted chemical activation or inhibition of kinases) and measured at a single time point. In this Technical Note, we present a new Integer Linear Programming (ILP) implementation of PHONEMeS, which is orders of magnitude more efficient and thus faster than the previous implementation. It also handles more complex data sets and experimental designs, thus broadening the application of PHONEMeS and the contextuality of the studies where it can be applied.

METHODS

PHONEMeS combines information from large-scale MS phosphoproteomic data with prior knowledge of signaling to contextualize signaling networks (Figure 1).

After phosphoproteomic measurements have been obtained from cell samples, the data are preprocessed and normalized. Then, differential analysis is performed to obtain the fold-change and significance of the regulation of quantified phosphosites. As a next step, a score is assigned to each measured phosphosite based on how much a measurement changes across the samples (based on fold-change and significance). Measurements with a significant change are classified as "Perturbed" (and they are assigned a negative score), whereas measurements with a low or not significant change are classified as "Control" (assigned a positive score) (Text S1). The lower a score, the more regulated a measurement is considered to be and vice versa for higher scores. The resulting PHONEMeS pathway solutions are expected to include as many "Perturbed" sites as possible while penalizing the inclusion of the "Control" sites.

We also built a PKN of signaling from large-scale pathway resources. It uses the meta-resource OmniPath by default,[5,6] but others can be used. 38 506 unique K/P−S relations are obtained from the OmniPath R package (September 2020 release). The way we generate the prior knowledge network from pathway resources is schematically depicted and described in Figure 1 (steps D and E in the caption) as well as in Text S1 ("Building the PKN" subsection). The set of interactions present in the PKN is used to subsequently train the PHONEMeS models. For that, we use ILP to map our inferred scores on the PKN and train our network. The purpose of the training is to identify a subset of mechanistic signaling interactions that can explain the data. Such interactions are directed and unsigned because we only aim at modeling the propagation of the signaling perturbation through the network of interacting proteins. By not requiring their sign, which is often unknown, we can include more phosphosites. In addition, the formulation is then simpler, and thus the computation is ore efficient. The resulting signaling models are then combined and used to generate testable hypotheses about signaling that can be subsequently validated. In the previous implementation of PHONEMeS (from here referred to as PHONEMeS-stoch) the stochastic approach used to train the networks was computationally demanding and required a cluster infrastructure to perform the analysis.

https://dx.doi.org/10.1021/acs.jproteome.0c00958
J. Proteome Res. 2021, 20, 2138−2144

The new ILP implementation (PHONEMeS-ILP) is much faster and can be readily used with a desktop computer, as illustrated later. We will use the term PHONEMeS to refer to the approach in general.

Additionally, PHONEMeS-ILP can now be used on various types of phosphoproteomic data sets obtained from various experimental scenarios (Text S1). In an experimental design where the effect of a perturbation is compared with a control, PHONEMeS-ILP (similar to PHONEMeS-stoch) can be used to infer affected signaling mechanisms downstream of the perturbed node, such as the target of a kinase inhibitor or stimulated cell receptors. Moreover, in a perturbation context, the phosphoproteomic data can be obtained at different time points to study the effects of the perturbation on the dynamics of cellular signaling across time. To infer at which time points the deregulations happen from time-series data, we have implemented the PHONEMeS-mult variant of PHONEMeS-ILP. This approach allows us to obtain a better dynamic overview of signaling propagation. Finally, phosphoproteomic data can also be obtained from nonperturbation experimental conditions, such as a case-control study comparing a disease to a healthy state. Here the challenge is that there is no specific molecular target of a perturbation. To handle this, we have implemented the so-called upside-down variant of PHONEMeS (PHONEMeS-ud). Here we use kinase activity estimation methods to identify kinases and phosphatases that are potentially deregulated and then use them as putative perturbed nodes to run PHONEMeS-ILP and infer signaling mechanisms upstream and downstream of those deregulated kinases and phosphatases. In this case, the training is performed in both ways: downstream of the targets first, and then upstream of the targets by flipping the PKN upside down.

PHONEMeS-ILP is available as a free open-source R-package. Depending on the application, the main functions are runPHONEMeS, runPHONEMeS_mult, and runPHONEMeS_UD. Within these functions, various parameters have been implemented to give the user better control of the quality and diversity of the solutions as well as the efficiency of the analysis. These parameters have been listed and described in the documentation of each function. Two solvers have been considered for PHONEMeS-ILP: CPLEX and CBC. CPLEX is more efficient and provides more extensive outputs but requires a license that can be obtained for free by academics. CBC is open source and freely available. Considering the important computational advantages, we currently support CPLEX, but we will monitor the consideration of CBC and other open-source alternatives for further developments and eventual use.

RESULTS

We provide here some examples of the different PHONEMeS variants.

Benchmarking the ILP Implementation of PHONEMeS

The PHONEMeS-ILP approach was initially applied in the breast cancer data set from ref 7 to model the effects of mTOR inhibition on signaling. This was also the same case study that was used to present PHONEMeS-stoch. Here we verified that the ILP implementation was able to capture the same signaling mechanisms that were reported in PHONEMeS-stoch (Text S2). The PHONEMeS models for these data were experimentally validated in ref 4, and they showed that mTOR inhibition had the effect of altering the activity of CDK2, an important kinase of cell-cycle regulation (Text S2). The major difference is that PHONEMeS-stoch is much more time-consuming. To illustrate the differences in computational times, we adapted the cluster scripts to make the parallel processes (50 nodes performing independent optimization for 50 generations) run in a sequence in a standalone machine (MacBook Pro with 2.9 GHz Intel Core i5 processor). For the benchmarking case study, the analysis took 8 h 45 min to run 50 generations with PHONEMeS-stoch, and we observed convergence in the solution after 10 generations (~2 h). In contrast, the ILP implementation was able to train the network with the CPLEX solver in less than 1 s on a desktop computer (0.07 s). Another difference to be noted and that is described in detail in the Supporting Information (Text S2) is that the deterministic nature of the ILP program allows PHONEMeS-ILP to converge into a global optimal solution that cannot be guaranteed by the stochastic algorithm implemented in PHONEMeS-stoch. These points illustrate how PHONEMeS-ILP is an efficient alternative to the PHONEMeS-stoch implementation.

Modeling of Endothelin Signaling with PHONEMeS-mult

The multiple-time-points variant of PHONEMeS-ILP (PHONEMeS-mult) was used to infer an endothelin receptor (EDNRB) regulatory signaling network based on prior knowledge and large-scale time-resolved phosphoproteomics for two types of melanoma cell lines.[8] For details, see Text S3. The network models of PHONEMeS-mult for this case study were experimentally validated and contributed to the system-level understanding of EDN signaling, which can subsequently be used for the selection of kinase inhibitors in combination with EDN treatments (Text S3).

Modeling of Aberrant Signaling in Colon Cancer with PHONEMeS-ud

The upside-down variant of PHONEMeS (PHONEMeS-ud) was applied to phosphoproteomic data collected from colon cancer cohorts (tumor and adjacent normal tissues).[9] The data were assembled as a single matrix with phosphosite measurements across both the tumor and normal tissues. Here we retained only those phosphosites that were quantified in at least 10% of the normal and tumor samples. This resulted in a table where 20 276 unique phosphosites were quantified across 197 samples (97 tumor and 100 healthy). We built the background network by using the OmniPath meta-resource, except for interactions coming from the ProtMapper source.[10] ProtMapper is based on text mining and increases coverage but can include false-positive kinase-to-substrate relations. We show here the results with the lower-coverage higher-curation network and include the results with the larger network to illustrate this trade-off in Text S5. From the 20 276 sites quantified, 2543 could be mapped to OmniPath (with ProtMapper-filtered interactions) as substrates of 544 unique protein kinases or phosphatases. Next, differential phosphorylation analysis was performed with Limma[11] with the false discovery rate (FDR) multiple testing correction. Of all of the measured sites, 987 were classified as significantly regulated or perturbed (with an FDR0.01 threshold value for significance and an absolute fold-change threshold value of abs(FC)2.5 up or down), 50 phosphosites were upregulated, and 937 were downregulated. From those 987 perturbed sites, 100 could be mapped in OmniPath. All 100 of these perturbed sites were also present in the PHONEMeS solution network. Moreover, on the basis of the estimated regulation levels of all of the

https://dx.doi.org/10.1021/acs.jproteome.0c00958
J. Proteome Res. 2021, 20, 2138−2144

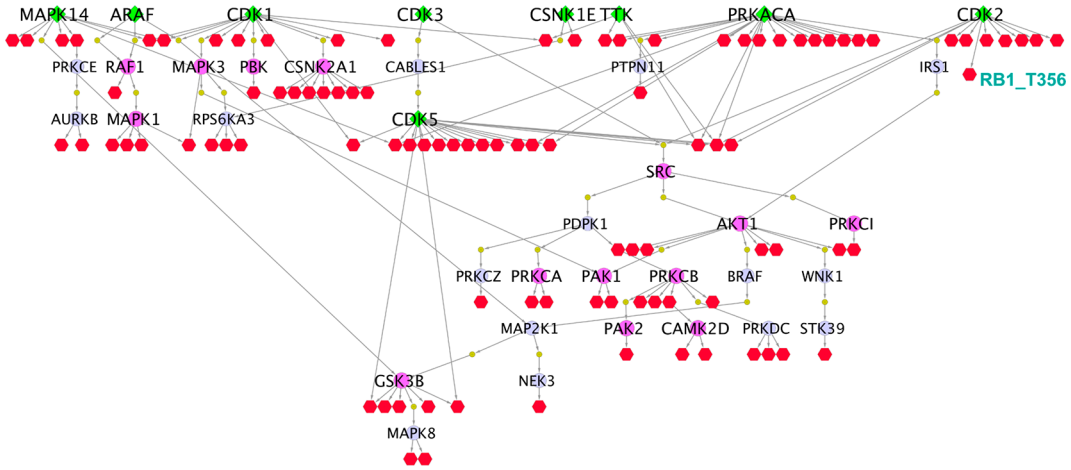


Figure 2. PHONEMeS-ud solution model of signaling for colon cancer cells. Target proteins (green diamonds) correspond to the highly regulated proteins (with absolute activity score higher than 10) as identified through the kinase activity estimation. Big purple circles correspond to intermediate signaling kinases inferred by PHONEMeS. Big pink circles are intermediate signaling kinases inferred by PHONEMeS but that were also predicted as regulated through kinase activity predictions (with an absolute activity score higher than 2 but smaller than 10). Small yellow circles correspond to possible phosphorylation sites of intermediate kinases (as inferred by PHONEMeS but that are not actually measured). Red hexagonal nodes correspond to measured phosphosites that were also included by PHONEMeS in the signaling model.

phosphosites, the activities of 517 unique kinases were estimated using a *z*-test.[12] Here the kinase activities were estimated based on the kinase−substrate relations coming from a collection of phosphoproteomic databases and text-mining tools.[10] Then, by setting a threshold on the inferred absolute kinase activity scores (>10), we identified 12 of the most regulated kinases, namely, CDC2, CDK2, CDK3, CK2A2, ARAF, ERN1, MAPK14, TTK, CSNK1E, PRKACA, CDK5, and CDK1. Of these, the CDC2 and CK2A2 were not present as protein kinases on the OmniPath set of kinase−substrate interactions and were excluded. ERN1, on the contrary, did not appear in the final solution as a result of filtering out ProtMapper interactions where the number of kinase−substrate interactions involving this protein went from 10 down to 2. The subsequent PHONEMeS-ud analysis yielded a contextualized perturbation network of signaling with 170 nodes and 189 interactions (Figure 2).

Whereas a detailed description of the signaling state of colon cancer goes beyond the purpose of this Technical Note, we performed some analyses to illustrate the insights that can be obtained from PHONEMeS. Most of the predicted active kinases, namely, MAPK14, PRKACA, CDK1, CDK2, CDK5, and TTK, are shown to be responsible for the regulation of a relatively large number of substrate proteins (7, 18, 9, 10, 17, 19, and 4, respectively). The rest (ARAF, CDK3, and CSNK1E) are shown to be responsible for the direct regulation of just one substrate at most (0, 0, and 1, respectively). This is due to the fact that these protein kinases were targeting a comparably fewer number of phosphorylation sites in the OmniPath background network than the first group (100 vs 700 on average). Furthermore, from the PHONEMeS solution, we see that most of the measured phosphosites in the network are terminal nodes, meaning that there are usually no interactions coming out of these nodes. However, there are also inferred signaling cascades that are mediated through the perturbed phosphosites. One such example is the PDPK1 → PRKCB_T500 → PRKCB → CAMK2D_T287 → CAMK2D cascade that is mediated by two significantly regulated phosphosites (namely, PRKCB_T500 and CAMK2D_T287). For this particular case, both of the enzyme-to-substrate relations involving these two measured phosphosites (PDPK1 → PRKCB_T500 and PRKCB → CAMK2D_T287) are phosphorylations, and we observe a decrease in phosphorylation (with log2 fold-change values of −2.335 and −0.689, respectively). Additionally, the abundance of other phosphosites on PDPK1, PRKCB, and CAMK2D target substrates decreases as well (with average log2 fold-change values of −2.335, −1.273, and −1.685, respectively). Another interesting observation that can be made is the potential activation of RB1_T356 from CDK2. In ref 9, the authors already proposed targeting the Rb1 tumor suppressor through CDK2 inhibition to repress cell proliferation. Finally, pathway enrichment analysis was performed over the kinases, which were predicted by PHONEMeS (meaning they are present in the model) with the fgsea method.[13,14] Unsurprisingly, KEGG_COLORECTAL_CANCER and GRADE_COLON_AND_RECTAL_CANCER_UP appeared among the top-ranking pathway sets (Figure S1a,b).

DISCUSSION

In this work, we have presented the most recent computational updates of our PHONEMeS-stoch method.[4] In particular, we have put a special emphasis on the integration of the ILP optimization techniques to provide information about the behavior of the biological systems representing signaling networks. As illustrated by different real-case applications, we are able to highlight pathway aberrations involved in diseases and to mechanistically understand how the deregulation of signaling affects the cellular functions. The results demonstrated the capability of PHONEMeS in providing useful signaling mechanistic descriptions and how they can be used to assist in the selection of combinations of drug treatments and to characterize deregulation in disease. PHONEMeS-ILP has been made publicly available as an R-package, and the examples indicated here have been included in the package. Full tutorials have been provided to reproduce the results shown in this Technical Note and help the users understand how to use PHONEMeS for their own data sets.

The new ILP formulation for PHONEMeS offered some important advantages. PHONEMeS-stoch relied on a Genetic

https://dx.doi.org/10.1021/acs.jproteome.0c00958
J. Proteome Res. 2021, 20, 2138−2144

Algorithm for the optimization and was generally computationally demanding. ILP implementation not only is much faster but also enables the method to be applied over alternative experimental designs besides just perturbation at one time point. These new capabilities include the option to contextualize signaling from time-course phosphoproteomics (PHONEMeS-mult) and data obtained from nonperturbation experimental setups (PHONEMeS-ud). To illustrate these improvements, we have described the application of PHONEMeS in three case studies: First, a benchmarking study was carried out using the breast cancer perturbation data set in which PHONEMeS-ILP was initially applied and introduced as a method. Here we showed that the ILP implementation was orders of magnitude faster and was able to reproduce the results of the original study, validating the new ILP formulation. It is to be noted, however, that the efficiency of PHONEMeS-ILP strongly depends on the input conditions and the parameter settings. Typically, the larger the number of experimental conditions considered for the analysis, the larger the amount of time it takes for PHONEMeS-ILP to reach the optimal solution. Optimization times in this case can typically go from subseconds, when considering one experimental condition, to a substantial number of minutes. This is to be expected because the number of perturbation targets normally increases for each additional experiment, and thus the number of paths connecting the perturbation targets to measurements also increases. Other parameters that can affect the computational time needed are the gap and populate parameters. The gap parameters can control the quality of the solutions we wish to retrieve, whereas the populate parameters can control the number of alternative solutions we wish to explore with PHONEMeS-ILP. By default, we set the gap parameters to 0, which forces PHONEMeS to run the analysis until the optimal guaranteed global solution has been reached. In addition, we typically use PHONEMeS to search and identify a large pool of alternative good solutions (5000 by default), and from that we retrieve the subset of the most diverse solutions (100 by default). Searching for alternative optimal solutions can substantially influence the runtime. In PHONEMeS, we can set any desired time limit through the time-limit parameter. Second, the application of the new PHONEMeS-mult (time-course form of PHONEMeS-ILP) for the modeling of endothelin signaling pathways was carried out.[8] The models were able to predict the presence of key kinases known to play a central role in cell migration, and major parts of the network proposed by PHONEMeS were experimentally validated.[8] Third, we applied PHONEMeS-ud to study the signaling state of colon cancer. In this case, PHONEMeS-ud was able to provide expected and potentially new mechanistic insights into colon cancer signaling. Our PHONEMeS-ud model can be used to propose causal mechanisms and potential crosstalk between signaling pathways. Whereas the network predictions for the first two case studies were experimentally validated, our study in colon cancer signaling requires further experimental work. The PHONEMeS-ILP method complements other recent network contextualization tools based on causal networks, including CellNOpt,[15] CARNIVAL,[16] and COSMOS.[17] These tools are used with other kinds of data, and while they can consider phosphoproteomic data, this is done indirectly by using kinase activities obtained from phosphoproteomics. Furthermore, PHONEMeS-ILP offers some distinct and complementary features to other similar tools, such as PHOTON,[2] TPS,[3] and the method from MacGilvray et al.,[18] as described in the Supporting Information (Text S4).

Despite its unique features, we note some limitations of PHONEMeS-ILP, in particular, limitations derived from being limited by the availability of prior knowledge: First, the prior knowledge might integrate some false-positive interactions. This is the case in the case study when including interactions obtained from ProtMapper, a resource based on text-mining. In that case, the PHONEMeS solution includes a few proteins that were neither kinases nor phosphatases (i.e., GAST and EIF4EBP1) despite the fact that we had only retained K/P−S interactions for building our prior knowledge background network. To address this issue, the ProtMapper interactions were filtered out before the PHONEMeS-ud analysis was performed, leading to a loss of coverage (Text S5). Another issue, as we noticed in the colon cancer application, was that only ∼12.5% of perturbed sites were present in OmniPath and hence could be potentially inferred in the solution network. In this way, we are losing ∼88% of the data and not exploiting the full potential of untargeted MS because PHONEMeS strictly infers regulatory interactions based on the prior knowledge. One way to alleviate this issue would be to expand the prior knowledge with alternative resources such as NetworKIN.[19] In these resources, the K/P−S relations are inferred based on consensus kinase recognition motifs and other information. This allows us to cover a wider range of phosphosites that can be targeted by enzymes, but these inferred interactions are typically less reliable. Additionally, there might be phosphosites for which we can not find an appropriate recognition motif; therefore, this approach will not guarantee that it can cover all of the data that can be measured. Along this line, as future work, the objective function of the new ILP formulation can be adapted to take into account the confidence about certain components in the prior knowledge and to prioritize certain interactions based on evidence from the literature or the functionality.[19] In summary, using different resources, larger networks and additional hypotheses can be generated, but these new insights are less trustable. There is a trade-off between curation and coverage that has to be carefully balanced in each case study.

Further future work includes expanding the types of prior knowledge and data beyond kinase-to-substrate interactions and phosphoproteomics, respectively. In particular, additional post-translational modifications[20] and conformational changes[21,22] provide information on the activity of proteins that complements phosphorylation. This will expand the applicability of PHONEMeS to systematically integrate large omic data sets with our current knowledge on molecular mechanisms. The resulting networks combine information from the data with prior knowledge in a single entity, a contextualized network of interactions, that is both human- and machine-readable. Such networks can, in turn, be used as inputs for machine-learning algorithms to identify the molecular patterns underlying specific phenotypes and to find disease biomarkers.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00958.

Text S1. PHONEMeS pipeline. Text S2. PHONEMeS-ILP consistency with original implementation. Text S3.

https://dx.doi.org/10.1021/acs.jproteome.0c00958
J. Proteome Res. 2021, 20, 2138−2144

Modeling of endothelin signaling with PHONEMeS-mult. Text S4. Comparison of PHONEMeS with other related tools. Text S5. Trade-offs of coverage and curation in the prior knowledge. Table S1. Set of variables used in the ILP formulation of PHONEMeS. Figure S1. Enrichment plots for PHONEMeS-ud model. Figure S2. Model solutions for the benchmark case study. Figure S3. Time-resolved network model of PHONEMeS-mult analysis (UACC257 cells). Figure S4. Time-resolved network model of PHONEMeS-mult analysis (A2058 cells). Figure S5. PHONEMeS-ud solution model (PDF)

AUTHOR INFORMATION

Corresponding Author

Julio Saez-Rodriguez − *Faculty of Medicine, Institute for Computational Biomedicine, Bioquant, INF267, Heidelberg University, 69120 Heidelberg, Germany; Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC−COMBINE), RWTH Aachen University, 52074 Aachen, Germany*; Email: julio.saez@uni-heidelberg.de

Authors

Enio Gjerga − *Faculty of Medicine, Institute for Computational Biomedicine, Bioquant, INF267, Heidelberg University, 69120 Heidelberg, Germany; Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC−COMBINE), RWTH Aachen University, 52074 Aachen, Germany;* orcid.org/0000-0001-8042-0395

Aurelien Dugourd − *Faculty of Medicine, Institute for Computational Biomedicine, Bioquant, INF267, Heidelberg University, 69120 Heidelberg, Germany; Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC−COMBINE), RWTH Aachen University, 52074 Aachen, Germany*

Luis Tobalina − *Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC−COMBINE), RWTH Aachen University, 52074 Aachen, Germany*

Abel Sousa − *European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), , Cambridge CB10 1SD, United Kingdom; Institute for Research and Innovation in Health (i3s), 4200-135 Porto, Portugal*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00958

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. E.G. formulated and implemented the ILP formulation of PHONEMeS. A.D. contributed to the implementation of PHONEMeS-ud as well as to the data preprocessing steps. L.T. implemented the CBC solver for PHONEMeS-ILP. A.S. performed the kinase enrichment analysis over the Colon Cancer data set. J.S.-R. conceived and supervised the project. E.G. wrote the manuscript, and J.S.-R., A.S., and A.D. revised the manuscript.

Notes

The authors declare the following competing financial interest(s): J.S.-R. has received funding from GSK and Sanofi and expects consultant fees from Travere Therapeutics.

PHONEMeS has been implemented as an R-package (https://github.com/saezlab/PHONEMeS-ILP) and is freely available under a GNU GPLv3 license, together with detailed documentation of its main functions and examples. Supplementary tables providing all of the network results for the examples mentioned in this manuscript are available at https://github.com/saezlab/Gjerga_etal_PHONEMeS_Supplement.

ACKNOWLEDGMENTS

The work was supported by the European Union's H2020 program (675585 Marie-Curie ITN "SymBioSys"); JRC for Computational Biomedicine, which is partially funded by Bayer; and the European Union Horizon 2020 research and innovation program under grant agreement no. 668858.

REFERENCES

(1) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717.

(2) Rudolph, J. D.; de Graauw, M.; van de Water, B.; Geiger, T.; Sharan, R. Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks. *Cell Syst* **2016**, *3* (6), 585−593.e3.

(3) Köksal, A. S.; Beck, K.; Cronin, D. R.; McKenna, A.; Camp, N. D.; Srivastava, S.; MacGilvray, M. E.; Bodík, R.; Wolf-Yadlin, A.; Fraenkel, E.; Fisher, J.; Gitter, A. Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data. *Cell Rep.* **2018**, *24* (13), 3607−3618.

(4) Terfve, C. D. A.; Wilkes, E. H.; Casado, P.; Cutillas, P. R.; Saez-Rodriguez, J. Large-Scale Models of Signal Propagation in Human Cells Derived from Discovery Phosphoproteomic Data. *Nat. Commun.* **2015**, *6*, 8033.

(5) Türei, D.; Korcsmáros, T.; Saez-Rodriguez, J. OmniPath: Guidelines and Gateway for Literature-Curated Signaling Pathway Resources. *Nat. Methods* **2016**, *13* (12), 966−967.

(6) Ceccarelli, F.; Turei, D.; Gabor, A.; Saez-Rodriguez, J. Bringing Data from Curated Pathway Resources to Cytoscape with OmniPath. *Bioinformatics* **2020**, *36* (8), 2632−2633.

(7) Wilkes, E. H.; Terfve, C.; Gribben, J. G.; Saez-Rodriguez, J.; Cutillas, P. R. Empirical Inference of Circuitry and Plasticity in a Kinase Signaling Network. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7719−7724.

(8) Schäfer, A.; Gjerga, E.; Welford, R. W.; Renz, I.; Lehembre, F.; Groenen, P. M.; Saez-Rodriguez, J.; Aebersold, R.; Gstaiger, M. Elucidating Essential Kinases of Endothelin Signalling by Logic Modelling of Phosphoproteomics Data. *Mol. Syst. Biol.* **2019**, *15* (8), e8828.

(9) Vasaikar, S.; Huang, C.; Wang, X.; Petyuk, V. A.; Savage, S. R.; Wen, B.; Dou, Y.; Zhang, Y.; Shi, Z.; Arshad, O. A.; et al. Clinical Proteomic Tumor Analysis Consortium. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **2019**, *177* (4), 1035−1049.e19.

(10) Bachman, J. A.; Gyori, B. M.; Sorger, P. K. Assembling a Phosphoproteomic Knowledge Base Using ProtMapper to Normalize Phosphosite Information from Databases and Text Mining. *bioRxiv* **2019**, 822668.

(11) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43* (7), e47.

(12) Hernandez-Armenta, C.; Ochoa, D.; Gonçalves, E.; Saez-Rodriguez, J.; Beltrao, P. Benchmarking Substrate-Based Kinase Activity Inference Using Phosphoproteomic Data. *Bioinformatics* **2017**, *33* (12), 1845−1851.

(13) Sergushichev, A. An Algorithm for Fast Preranked Gene Set Enrichment Analysis Using Cumulative Statistic Calculation. *bioRxiv* **2016**, 060012 DOI: 10.1101/060012v1.
(14) Korotkevich, G.; Sukhov, V.; Budin, N.; Shpak, B.; Artyomov, M. N.; Sergushichev, A. Fast Gene Set Enrichment Analysis. *bioRxiv* **2021**, 060012 DOI: 10.1101/060012v3.
(15) Gjerga, E.; Trairatphisan, P.; Gabor, A.; Koch, H.; Chevalier, C.; Ceccarelli, F.; Dugourd, A.; Mitsos, A.; Saez-Rodriguez, J. Converting Networks to Predictive Logic Models from Perturbation Signalling Data with CellNOpt. *Bioinformatics* **2020**, *36* (16), 4523−4524.
(16) Liu, A.; Trairatphisan, P.; Gjerga, E.; Didangelos, A.; Barratt, J.; Saez-Rodriguez, J. From Expression Footprints to Causal Pathways: Contextualizing Large Signaling Networks with CARNIVAL. *NPJ. Syst. Biol. Appl.* **2019**, *5*, 40.
(17) Dugourd, A.; Kuppe, C.; Sciacovelli, M.; Gjerga, E.; Gabor, A.; Emdal, K. B.; Vieira, V.; Bekker-Jensen, D. B.; Kranz, J.; Bindels, E. M. J.; et al. Causal Integration of Multi-Omics Data with Prior Knowledge to Generate Mechanistic Hypotheses. *Mol. Syst. Biol.* **2021**, *17*, e9730.
(18) MacGilvray, M. E.; Shishkova, E.; Chasman, D.; Place, M.; Gitter, A.; Coon, J. J.; Gasch, A. P. Network Inference Reveals Novel Connections in Pathways Regulating Growth and Defense in the Yeast Salt Response. *PLoS Comput. Biol.* **2018**, *13* (5), e1006088.
(19) Linding, R.; Jensen, L. J.; Ostheimer, G. J.; van Vugt, M. A. T. M.; Jorgensen, C.; Miron, I. M.; Diella, F.; Colwill, K.; Taylor, L.; Elder, K.; et al. Systematic Discovery of in Vivo Phosphorylation Networks. *Cell* **2007**, *129* (7), 1415−1426.
(20) Ochoa, D.; Jarnuczak, A. F.; Viéitez, C.; Gehre, M.; Soucheray, M.; Mateus, A.; Kleefeldt, A. A.; Hill, A.; Garcia-Alonso, L.; Stein, F.; et al. The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **2020**, *38* (3), 365−373.
(21) Thygesen, C.; Boll, I.; Finsen, B.; Modzel, M.; Larsen, M. R. Characterizing disease-associated changes in post-translational modifications by mass-spectrometry. *Expert Rev. Proteomics* **2018**, *15* (3), 245−258.
(22) Cappelletti, V.; Hauser, T.; Piazza, I.; Pepelnjak, M.; Malinovska, L.; Fuhrer, T.; Li, Y.; Dorig, C.; Boersema, P.; Gillet, L.; et al. Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **2021**, *184* (2), 545−559.e22.

https://dx.doi.org/10.1021/acs.jproteome.0c00958
J. Proteome Res. 2021, 20, 2138−2144