

piNET: a versatile web platform for downstream analysis and visualization of proteomics data

Behrouz Shamsaei¹, Szymon Chojnacki¹, Marcin Pilarczyk¹, Mehdi Najafabadi¹, Wen Niu¹, Chuming Chen², Karen Ross³, Andrea Matlock⁴, Jeremy Muhlich⁵, Somchai Chutipongtanate^{6,7}, Jie Zheng⁸, John Turner⁹, Dušica Vidović⁹, Jake Jaffe¹⁰, Michael MacCoss¹¹, Cathy Wu^{2,3}, Ajay Pillai¹², Avi Ma'ayan¹³, Stephan Schürer⁹, Michal Kouril¹⁴, Mario Medvedovic^{1,15} and Jarek Meller^{1,14,16,*}

¹Department of Environmental and Public Health Sciences, University of Cincinnati College of Medicine, USA, ²Center for Bioinformatics & Computational Biology; University of Delaware, USA, ³Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, USA, ⁴Department of Biomedical Sciences, Cedars-Sinai Medical Center, USA, ⁵Department of Systems Biology, Harvard Medical School, USA, ⁶Department of Cancer Biology, University of Cincinnati College of Medicine, USA, ⁷Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Thailand, ⁸Department of Genetics, University of Pennsylvania Perelman School of Medicine, USA, ⁹Department of Pharmacology, Miller School of Medicine, Sylvester Comprehensive Cancer Center, Center for Computational Science, University of Miami, Miami, USA, ¹⁰Broad Institute of MIT and Harvard & Inzen Therapeutics, USA, ¹¹Department of Genome Sciences, University of Washington, USA, ¹²Human Genome Research Institute, National Institutes of Health, Bethesda, USA, ¹³Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, USA, ¹⁴Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, USA, ¹⁵Department of Biomedical Informatics, University of Cincinnati College of Medicine, USA and ¹⁶Department of Electrical Engineering and Computer Science, University of Cincinnati, USA

Received March 25, 2020; Revised April 29, 2020; Editorial Decision May 08, 2020; Accepted May 27, 2020

ABSTRACT

Rapid progress in proteomics and large-scale profiling of biological systems at the protein level necessitates the continued development of efficient computational tools for the analysis and interpretation of proteomics data. Here, we present the piNET server that facilitates integrated annotation, analysis and visualization of quantitative proteomics data, with emphasis on PTM networks and integration with the LINCS library of chemical and genetic perturbation signatures in order to provide further mechanistic and functional insights. The primary input for the server consists of a set of peptides or proteins, optionally with PTM sites, and their corresponding abundance values. Several interconnected workflows can be used to generate: (i) interactive graphs and tables providing comprehensive annotation and mapping between peptides and proteins with PTM sites; (ii) high resolution and interactive visualization for enzyme-substrate networks, including

kinases and their phospho-peptide targets; (iii) mapping and visualization of LINCS signature connectivity for chemical inhibitors or genetic knockdown of enzymes upstream of their target PTM sites. piNET has been built using a modular Spring-Boot JAVA platform as a fast, versatile and easy to use tool. The Apache Lucene indexing is used for fast mapping of peptides into UniProt entries for the human, mouse and other commonly used model organism proteomes. PTM-centric network analyses combine PhosphoSitePlus, iPTMnet and SIGNOR databases of validated enzyme-substrate relationships, for kinase networks augmented by DeepPhos predictions and sequence-based mapping of PhosphoSitePlus consensus motifs. Concordant LINCS signatures are mapped using iLINCS. For each workflow, a RESTful API counterpart can be used to generate the results programmatically in the json format. The server is available at <http://pinet-server.org>, and it is free and open to all users without login requirement.

*To whom correspondence should be addressed. Tel: +1 513 558 1958; Email: mellerj@ucmail.uc.edu

INTRODUCTION

Methods and tools for computational proteomics data analysis evolve constantly in order to match rapid advances in proteomics and enabled by them large scale efforts aiming to profile biological systems at the protein level (1–3). In this context, many methods and data processing tools have been developed for the identification and quantification of peptides and proteins from biological samples using mass spectrometry-based approaches (4–6). Further downstream analysis is subsequently performed to facilitate biological interpretation of such obtained quantitative results, with its own distinct set of challenges (6), and an interdependent ecosystem of databases, tools and resources for proteomics research (7).

Increasingly, proteomic profiling efforts involve identification and quantification of specific proteoforms that may exhibit distinct activities and functions, including those resulting from protein post-translational modifications (PTMs) (8). The importance of protein phosphorylation, acetylation, methylation and other PTMs that are involved in essential cellular processes has led to the development of databases and tailored resources, such as PhosphoSitePlus (9), Phospho.ELM (10) and PhosphoPep (11) for phospho-proteomics, as well as comprehensive databases of PTMs and aggregators of related functional annotations, such as iPTMnet (12), dbPTM (13) or SIGNOR (14).

This contribution introduces a web server for both interactive and programmatic analysis and visualization of peptide and protein level proteomic data, with special emphasis on PTMs. The new tool, dubbed piNET, aims to facilitate mapping, annotation and analysis of post-translational modification sites, as well as modifying enzymes that target these sites in the context of biological pathways. As illustrated in Figure 1, piNET integrates iPTMnet (Protein Information Resource), PhosphoSitePlus, SIGNOR and other proteomic resources with PTM network analysis, while making use of a fast approach for peptide mapping and a custom D3 library for versatile visualization.

Furthermore, piNET connects proteomics profiles with transcriptional and proteomics signatures generated by The Library of Integrated Network-based Cellular Signatures (LINCS) project (15), which aims to systematically collect Omics signatures of genetic and chemical perturbations, including signatures induced by genetic knock-downs or small molecule inhibitors of the modifying enzymes, such as kinases. By seamless integration with Enrichr (16,17), iLINCS (<http://ilincs.org>) and related tools that facilitate interaction with the LINCS library of signatures, one can apply the Connectivity Map approach (18) in order to gain further mechanistic insights into signalling cascades potentially driving the underlying biological states and their proteomics signatures.

MATERIALS AND METHODS

piNET has been implemented as a modular web server, using the Spring-Boot JAVA platform in conjunction with state-of-the-art D3 JavaScript libraries and their in-house developed extensions for highly customizable visualization. As a result, piNET can be executed within the browser

and does not require installing additional plugins or applications. Programmatic access to piNET workflows is enabled through the corresponding RESTful Application Programming Interface (API) methods. Documentation and help pages illustrate piNET workflows and the use of APIs with realistic examples, including Python scripts that invoke these respective APIs and embed the results in the json format into other applications.

piNET integrates several databases and resources for proteomics studies (see Table 1) in order to streamline mapping, annotation and analysis of post-translational modification sites, as well as modifying enzymes that target these sites. Multiple definitions of PTMs can be used interchangeably, including shorthand notation (e.g., pS, aK), mass difference (DeltaMass) identified experimentally, or PTM ontologies, including PsiMOD (<https://www.ebi.ac.uk/ols/ontologies/mod>) and UniMOD (<http://www.unimod.org>). As part of the initial processing of input peptide/PTM lists, using a Pride-MOD utility (19) and in-house developed translation modules, PTMs are mapped into PsiMOD and UniMOD entries. Such unified encoding can be used, in conjunction with mapping into proteins, as a basis for generating unique proteoform identifiers (provisionally implemented in the form of Protein Line Notation tokens).

piNET can be used to query UniProt or Prosite APIs in order to map peptides and PTM sites into proteins from well annotated and continuously updated proteomes available through these authoritative resources (20–22). However, for fast mapping of large sets of peptides and PTMs, a locally indexed version of the UniProtKB/Swiss-Prot and TrEMBL protein sequence databases for a number of commonly used organisms, including *homo sapiens* and *mus musculus* (please refer to piNET for a complete list of organisms), has also been implemented using the state-of-the-art Apache Lucene indexing by Chen *et al.* (23). Using this fast peptide matching approach, an average search time of 0.02 s per peptide has been observed in our tests for a random sample of 20 000 peptides of length 10.

In addition, in order to overcome the limitations on the total number of peptides that can be submitted directly due to the maximum length of the URL, a file upload mode can be used to submit large sets of peptides and PTMs (with their abundance values), using either the MaxQuant evidence file or a generic multi-column csv format. This option also enables submission of datasets with proteomic profiles of multiple samples and meta-data descriptors to define groups of samples, with the goal of deriving a protein or peptide level differential expression signature for further analysis. Volcano plots can be generated in this workflow to guide the choice of differentially expressed peptides/PTMs/protein moieties based on the analysis of the effect size (fold change) and statistical significance (t-test based *P*-values).

Furthermore, piNET mines iPTMnet, PhosphoSitePlus, and SIGNOR for functional annotations of PTMs, and integrates these annotations to identify kinases and other modifying enzymes that are experimentally known to target PTM sites involved, and as a basis for PTM network analysis and visualization. For kinases, in addition to known modifier-PTM site pairs, a well performing method developed recently using deep learning, DeepPhos (24) is used

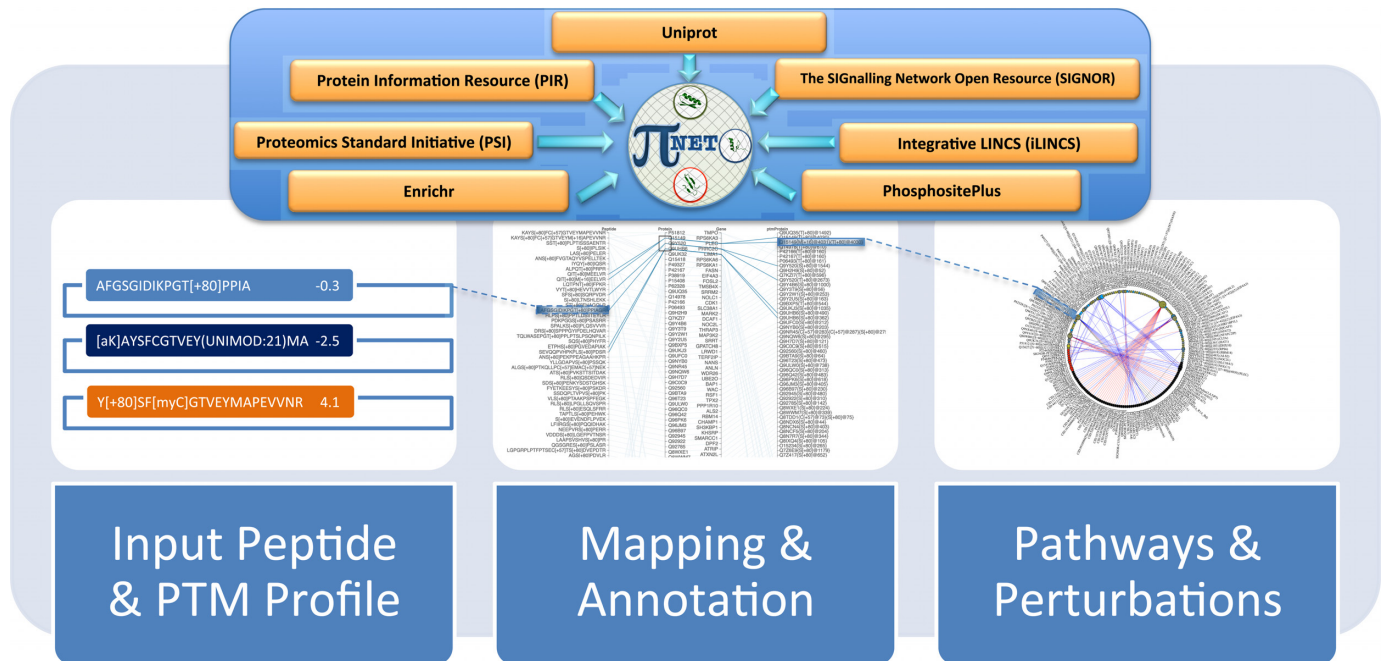


Figure 1. The overall flowchart of piNET proteoform-centric mapping, annotation and visualization workflow.

Table 1. The list of databases and resources integrated in piNET with the goal of providing an easy to use and intuitive interface for the analysis and visualization of proteomics data. Plus signs are used to indicate components available as part of each of the resources listed

Name	Web site	PTM mapping & annotations	Pathways & perturbations
UniProt	https://www.uniprot.org	+	
PsiMOD	https://www.psidev.info/MOD	+	
PROSITE	https://prosite.expasy.org	+	
PhosphoSitePlus	https://www.phosphosite.org/	+	+
iPTMnet	http://proteininformationresource.org/iPTMnet	+	+
Signor	http://signor.uniroma2.it	+	+
DeepPhos	https://github.com/USTC-HIlab/DeepPhos	+	+
iLINC	https://ilincs.org/		+
Enrichr	https://amp.pharm.mssm.edu/Enrichr		+
Reactome	https://reactome.org		+

to generate predicted kinase-PTM site pairs. These predictions are further augmented by sequence-based mapping of sites that share a high degree of motif similarity with known target sites, using PhosphoSitePlus known target sequences and consensus motif definition. piNET modular annotation and visualization engines can be extended to include other representative methods for the prediction of kinase-substrate pairs and related resources (25–29), such as NetworKIN (27), in order to further expand kinome annotations.

Finally, piNET connects protein level signatures with transcriptional and proteomics signatures generated by the LINC project (30). A simple signature consisting of a set of up- and down-regulated proteins can be first compared with a large set of LINC mRNA level signatures as readily available proxies by utilizing the aforementioned Enrichr APIs to generate the results of enrichment analysis, with emphasis on gene sets consisting of gene up- and/or down-regulated in response to kinase loss or gain of function, and thus representing kinase gain or loss of function signatures. Furthermore, piNET uses iLINC APIs to connect protein level relative or differential expression profiles (differential

abundance vectors) either with LINC mRNA level signatures as a proxy, or directly with proteomics signatures generated by the LINC projects, including those induced by the loss or gain of function of relevant modifying enzymes (when available). It should be stressed, however, that the overall number of proteomics signatures generated by the LINC project is much smaller compared to transcriptional profiling (30).

RESULTS

piNET provides an intuitive and easy to use interface, divided into modular workflows. Several use cases, including those stemming from the LINC project, are incorporated as examples in order to illustrate piNET functionality, network-based visualization and integration with biological domain knowledge. We are also using these use cases to illustrate potential pitfalls, such as those related to the problem of projecting proteoform level data onto gene level in order to facilitate pathway and other functional analyses. Strategies to overcome these limitations are also discussed.

Peptide to protein workflow

As the first step in downstream analysis, piNET can be used to annotate, map and analyze a set of peptide moieties, including those with post-translational modifications. Starting from a set of peptides, and optional modifications within those peptides, piNET provides mapping into both canonical and other isoforms included either in the UniProtKB/Swiss-Prot and/or TrEMBL databases. piNET also provides mapping and harmonization of PTM meta-data, using PsiMOD and UniMOD ontologies, and generating Protein Line Notation tokens to represent each peptide moiety with modifications for automated annotation exchange systems. Mapping of peptides and PTM sites into proteins and genes is summarized as intuitive interactive graphs and tables that can be downloaded, or alternatively generated using API calls for integration with other resources. An example of a graphical summary of peptide to protein to PTM mapping is shown in Figure 2 for a subset of P100 phospho-peptides (31) that are being used to profile responses to cellular perturbation by the LINCS project. Submissions and updates in the first workflow (peptide to protein tab) are automatically propagated to the subsequent workflows, while offering several options for aggregating peptide level data into protein level abundance data, including averaging peptide abundance values for peptides mapping into a protein, while ignoring peptides mapping into multiple proteins.

PTM to modifying enzyme workflow

This PTM-centric workflow aims to elucidate signaling cascades converging on PTM sites by mapping known and predicted upstream enzymes that target those PTM sites. An interactive and high resolution PTM-centric network views are provided by embedding PhosphoSitePlus, iPTMnet and SIGNOR databases of enzyme-substrate relationships. For phosphorylation, known kinase - target phospho sites are derived from PhosphoSitePlus (9) to provide a site specific PTM-centric kinome network view, and shed light onto phosphorylation cascades that may be involved. On the other hand, iPTMnet provides comprehensive mapping and annotation of both kinase—phosphopeptide pairs (which largely overlaps with that of PhosphoSitePlus), as well as other modifying enzymes and their targets (12). Finally, SIGNOR can be used in piNET to map carefully curated causal relationships, available for over 2800 human proteins participating in signal transduction. Specifically, PTMs causing a change in protein concentration or activity have been curated and linked to the modifying enzymes (14).

The enzyme-substrate relations can be visualized using highly customizable interactive graphs that capture biological entities and relationships potentially associated with the observed patterns in protein expression profiles. For example, the casual relationships extracted from SIGNOR can be represented as edges in the graph, using the color of an edge to indicate the activation/inactivation relationships between signalling entities. A bipartite graph view for P100 kinome network, using SIGNOR causal relationships, is illustrated in Figure 3. The moieties that are being quantified in a proteomic assay (here P100 phospho-peptides with

their Z-score transformed normalized abundance values) are shown as nodes in the right column, while the modifying enzymes that target these PTMs in the left column. When quantitative peptide or protein abundance data are given, the abundance levels of peptide/PTM/protein moieties are represented by the color of a node, with yellow corresponding to high relative abundance and blue to low relative abundance, respectively. Such views provide a visual representation of relationships between peptide/PTM/protein moieties and their abundance profiles, and other relevant biological entities or processes.

Protein to pathway & perturbation workflow

A set of peptides (with PTMs) and the associated (differential) abundance values can be projected onto the protein (and thus gene) level for the purpose of further pathway analysis and visualization. Peptides matching multiple proteins can be ignored at that stage, while the geometric average is used to assign per protein value for multiple peptides matching the same protein. However, several other options to map peptide level abundance values into proteins are also available. In addition, the user can directly specify per protein values when submitting protein level profiles in the Protein2Pathway tab. Subsequently, piNET can be used to facilitate pathway enrichment analysis with Enrichr, and to perform further mapping into Reactome pathways for interactive exploration, while generating high resolution visualization and tabulated results.

Since PTMs mediate cell growth, death and differentiation through intracellular signal transduction cascades and the resulting transcriptional changes, gene and protein expression signatures and the signature connectivity analysis can potentially be used to derive useful information and mechanistic insights. LINCS provides a large library of transcriptional and proteomic signatures measured in response to molecular, genetic and disease perturbations, including the loss of function (due to genetic KDs and small molecule inhibitors) of modifying enzymes. To capitalize on this resource, piNET can be used to analyze protein level data in conjunction with LINCS signatures of cellular perturbations, by using iLINCS APIs or exploring the results interactively in iLINCS (<http://ilincs.org>). An example of connectivity analysis for a user provided input L1000 signature (of a multi-targeted kinase inhibitor midostaurin) is shown in Figure 4, with other strongly concordant or discordant LINCS signatures connected by edges in the circos plot. Note that other related kinase inhibitors are identified, suggesting overlapping kinase targets.

DISCUSSION

Downstream analysis and visualization of MS-based proteomic data is essential for biological interpretation. However, this process poses challenges to most biologists. Several excellent software tools to facilitate downstream analysis have been developed by the proteomic community, including Perseus (31), PeptideShaker (32) and Scaffold (www.proteomesoftware.com). However, these are standalone software packages that require installation (and for Scaffold, purchasing a licence key), and have dependencies

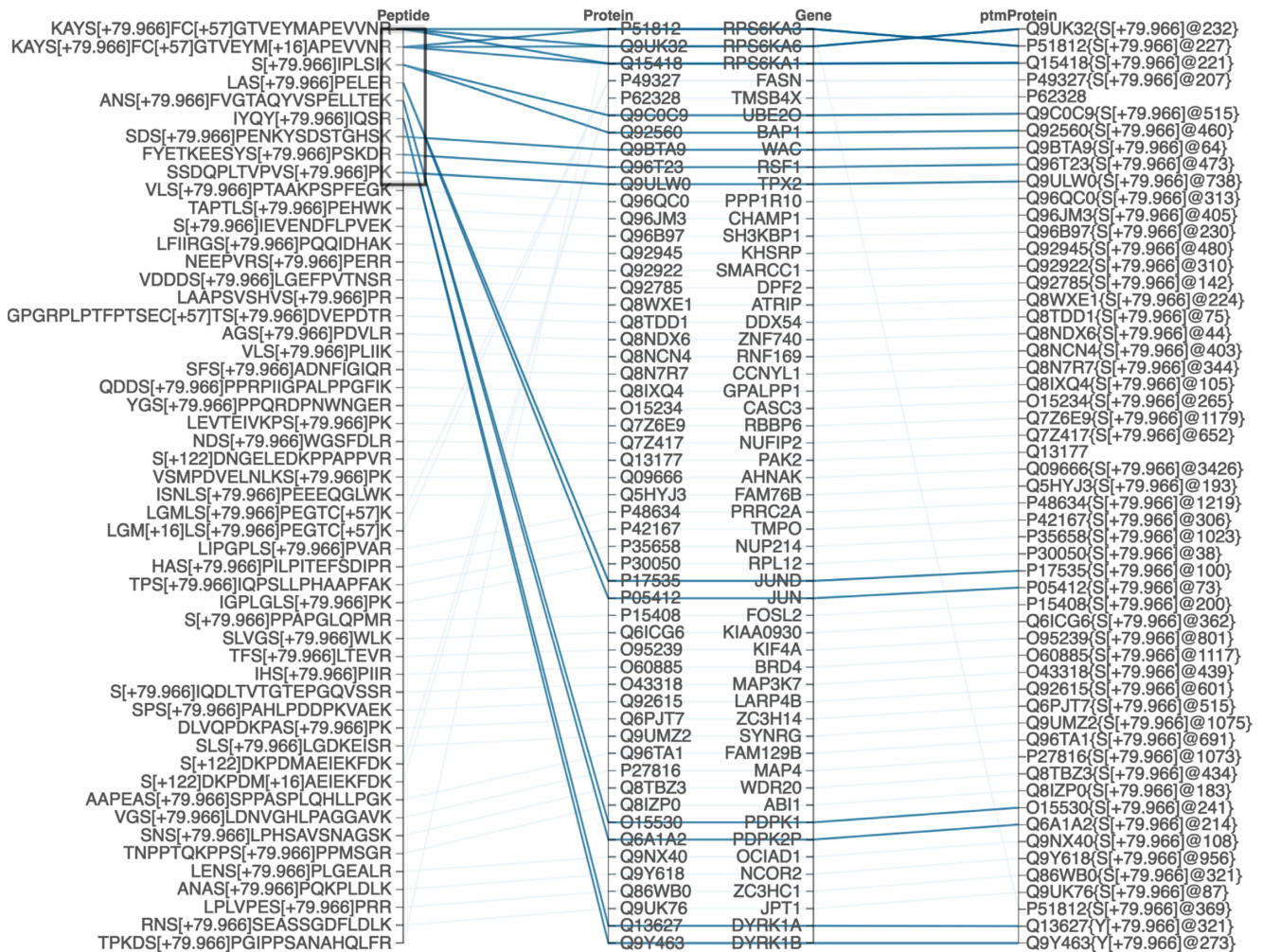


Figure 2. Peptide to protein to gene mapping for a subset of P100 representative phospho-peptides with PTM modification sites mapped into the corresponding protein and gene entities. Note that some peptides match multiple proteins and/or genes (many to many mapping), as indicated by thick blue edges for the highlighted peptides.

(e.g. on specific libraries) that can represent a significant barrier for adoption among biologists. Importantly, these stand-alone tools do not provide a programmatic interface for integration with other resources, and instead need to be integrated with local data processing pipelines.

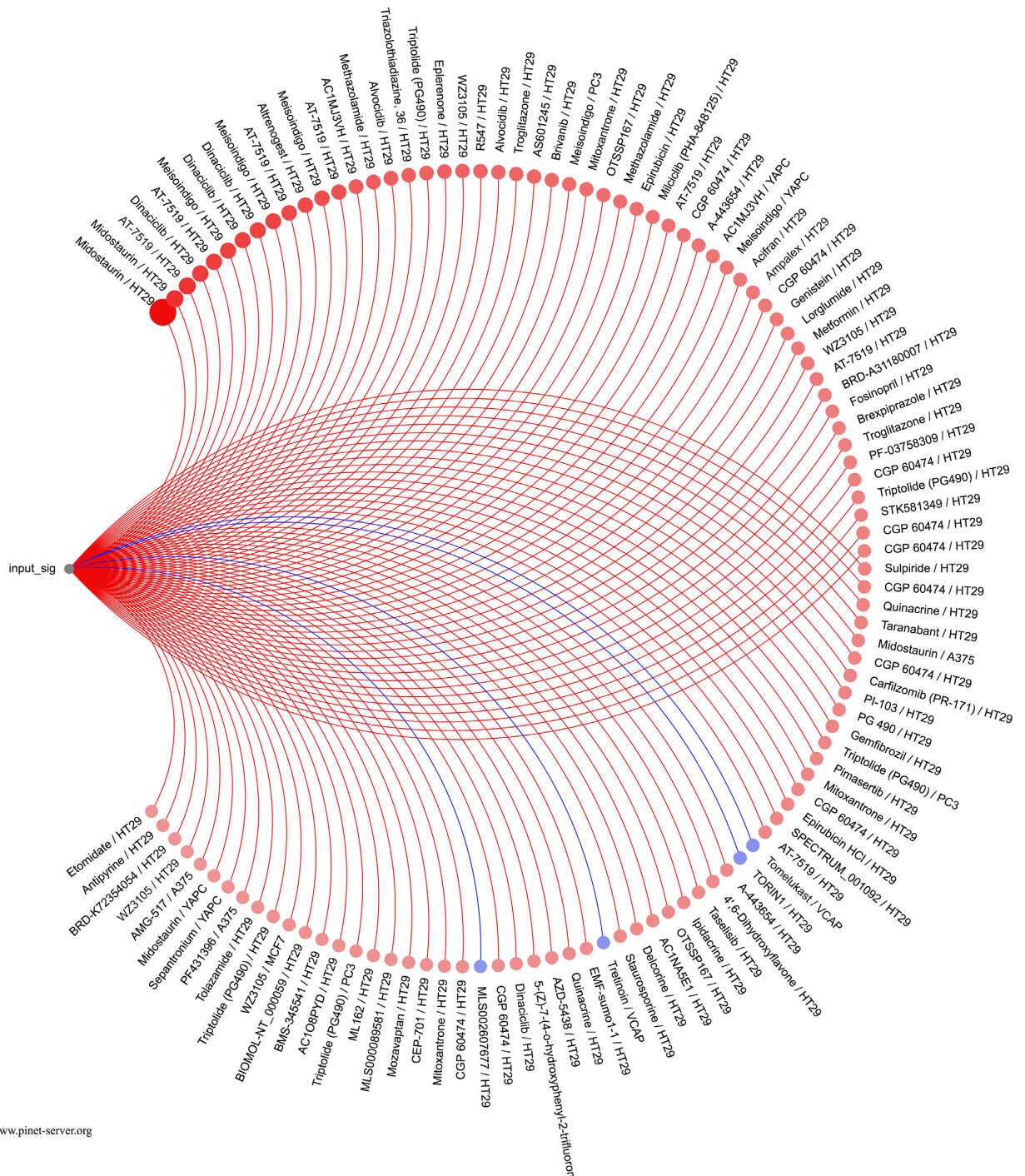
piNET has been designed to provide an integrated web platform for analysis, interpretation and visualization of large-scale proteomics data. To that end, piNET enables fast peptide/PTM to protein mapping, harmonization of meta-data pertaining to PTMs, systematic PTM to modifying enzyme mapping and other functional annotations, coupled with high quality visualization of PTM networks and protein pathways. We believe that piNET adds significantly to the ecosystem of tools for downstream proteomic data analysis by integrating these individual components and annotation resources, by coupling them with a high quality visualization engine, and by making annotation and analysis workflows available as API methods for easy integration with other tools and resources for proteomics. To the best of our knowledge, there are no web-based tools that enable fast, large-scale mapping of peptides and PTMs, integrated

with subsequent PTM network and signature connectivity analyses for biological insights. However, we would like to recognize another peptide-centric web tool for MS-based proteomics data analysis, namely Pathway Palette (33) that uses a client-server architecture to provide mapping from peptides to biological pathways.

Since protein pathway activation (either through a higher protein expression or PTM, for instance) often leads to transcriptional and proteomic signatures downstream from signalling cascades, one can use connectivity map approach (18) to provide additional mechanistic insights, e.g. by identifying likely modifying enzymes with concordant gain of function or discordant loss of function signatures. piNET can be used to connect user provided (or computed on-the-fly from a quantitative proteomic dataset) relative or differential protein expression signatures with the LINCS library of chemical and genetic perturbation signatures, as illustrated in Figure 4. However, several inherent limitations of this approach must also be considered, including the fact that most of the LINCS perturbation profiling has been in fact performed at the transcriptional level, using



Figure 3. Functional annotation and visualization of a PTM network for a set of P100 phospho-peptide probes using curated causal annotations from SIGNOR, overlaid with a representative P100 expression (abundance) profile for a non-specific kinase inhibitor, staurosporin, generated as part of the LINC project. The Z-score transformed relative abundance of P100 phospho-peptides in the right column is indicated by the color of nodes: blue for lower (down-regulated sites) and yellow for higher (up-regulated sites) than average, respectively. Note that P100 phospho-sites and their known modifiers (or regulators) are connected using red or blue edges for positive (or activating) versus negative (or inactivating) relationships, respectively. The modifying enzymes (kinases) in the left column are shown as red nodes, whereas other type of regulators, such as those involved in protein-protein interactions, are shown as black nodes.



PiNET-server @ www.pinet-server.org

Figure 4. Connectivity analysis for a user provided input signature (here L1000 signature of a multi-targeted kinase inhibitor midostaurin), represented by the gray node in the figure. Strongly connected LINCS signatures of chemical perturbations, including midostaurin in different cell lines and related drugs, are shown in red for positive and blue for negative correlations, respectively. For each perturbation, the name of the compound and the cell line used are indicated in the labels. The shade of the node represents the strength of correlation (darker shades corresponding to higher correlations squared), whereas the size of the node represents its statistical significance (*P*-values computed by iLINCS).

the L1000 landmark gene set, and to a smaller degree, at the proteomics level using targeted assays, such as P100 phospho-proteomic and GCP global chromatin profiling (30,34). Consequently, the relevant perturbations may not have been profiled at the protein level, or the overlap between peptide and protein moieties from user defined and LINCS proteomics profiles may be too small for reliable connectivity analysis. At the same time, the utility of using mRNA levels as a proxy for proteomics signatures can be limited. Although widely used when considering which pathways or biological processes could be involved, this step constitutes an approximation, especially in the case of PTMs. Therefore, proper care needs to be applied to interpret the results.

In this regard, we would like to comment that the LINCS and related data generation efforts are on-going, and we expect the number of signatures available for such analyses to grow over time. We also aim to further integrate piNET and its visualization engine with the recently published PTM-sigDB that enables PTM-Signature Enrichment Analysis (PTM-SEA) of phosphoprotein signature data (35), and thus should alleviate some of the above-mentioned limitations as well.

ACKNOWLEDGEMENTS

We would like to thank the developers of tools and resources for proteomics integrated into piNET, for making their software packages, web servers and API methods available to the community. We would like to acknowledge our early users, especially Drs. Rob McCullumsmith, Ken Greis, Michael Wagner and their groups, members of the BD2K-LINCS Data Coordination and Integration Center, as well as LINCS consortium participants, for their feedback and encouragement.

FUNDING

National Institutes of Health [U54 HL127624, P30 ES006096, R01 MH107487, 1T32CA236764, UL1TR001425, U01GM120953, in part]. Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

- Gillet, L.C., Leitner, A. and Aebersold, R. (2016) Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. *Annu. Rev. Analyt. Chem.*, **9**, 449–472.
- Ebhardt, H.A., Root, A., Sander, C. and Aebersold, R. (2015) Applications of targeted proteomics in systems biology and translational medicine. *Proteomics*, **15**, 9193–9208.
- Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature*, **537**, 347–355.
- Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.*, **11**, 2301–2319.
- Bilbao, A., Varesio, E., Luban, J., Strambio-De-Castillia, C., Hopfgartner, G., Müller, M. and Lisacek, F. (2015) Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics*, **15**, 964–980.
- Sinitcyn, P., Rudolph, J. and Cox, J. (2018) Computational methods for understanding mass spectrometry-based shotgun proteomics data. *Annu. Rev. Biomed. Data Sci.*, **1**, 207–234.
- Perez-Riverol, Y., Alpi, E., Wang, R., Hermjakob, H. and Vizcaino, J.A. (2015) Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics*, **15**, 930–950.
- Smith, L. and Kelleher, N. (2013) Proteoform: a single term describing protein complexity. *Nat. Methods*, **10**, 186–187.
- Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B. and Gnäd, F. (2019) 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.*, **47**, D433–D441.
- Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J. and Diella, F. (2011) Phospho.ELM: a database of phosphorylation sites – update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
- Bodenmiller, B., Campbell, D., Gerrits, B., Lam, H., Jovanovic, M., Picotti, P., Schlapbach, R. and Aebersold, R. (2008) PhosphoPep – a database of protein phosphorylation sites in model organisms. *Nat. Biotechnol.*, **26**, 1339–1340.
- Huang, H., Arighi, C.N., Ross, K.E., Ren, J., Li, G., Chen, S.-C., Wang, Q., Cowart, J., Vijay-Shanker, K. and Wu, C.H. (2018) iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res.*, **46**, D542–D550.
- Huang, K.Y., Lee, T.Y., Kao, H.J., Ma, C.T., Lee, C.C., Lin, T.H., Chang, W.C. and Huang, H.D. (2019) dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic Acids Res.*, **47**, D298–D308.
- Perfetto, L., Briganti, L., Calderone, A., Cerquone Perpetuini, A., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., Pavlidou, T. et al. (2016) SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.*, **44**, D548–D554.
- Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A.B., Silverstein, M.C., Lachmann, A. et al. (2018) The library of integrated Network-Based cellular signatures NIH Program: System-Level cataloging of human cells response to perturbations. *Cell Syst.*, **6**, 13–24.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma'ayan, A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. et al. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929–1935.
- Yasset, P.-R., Uszkoreit, J., Sanchez, A., Ternent, T., del Toro, N., Hermjakob, H., Vizcaino, J.A. and Wang, R. (2015) ms-data-core-api: an open-source, metadata-oriented library for computational proteomics. *Bioinformatics*, **31**, 2903–2905.
- The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuče, B.A., Hulo, N., Bridge, A., Bougueleret, L. and Xenarios, I. (2013) New and continuing developments at PROSITE. *Nucleic Acids Res.*, **41**: D344–D347.
- de Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.*, **34**, W362–W365.
- Chen, C., Li, Z., Huang, H., Suzek, B.E., Wu, C.H. and UniProt Consortium. (2013) A fast peptide match service for UniProt Knowledgebase. *Bioinformatics*, **29**, 2808–2809.
- Luo, F., Wang, M., Liu, Y., Zhao, X.M. and Li, A. (2019) DeepPhos: prediction of protein phosphorylation sites with deep learning. *Bioinformatics*, **35**, 2766–2773.
- Gnäd, F., Gunawardena, J. and Mann, M. (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.*, **39**, D253–D260.
- Oughtred, R., Stark, C., Breitkreutz, B.J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R. et al.

- (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.
27. Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J., Linding, R. *et al.* (2014) KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods*, **11**, 603–604.
 28. Lee, T.Y., Bo-kai Hsu, J., Chang, W.C. and Huang, H.D. (2011) RegPhos: a system to explore the protein kinase-substrate phosphorylation network in humans. *Nucleic Acids Res.*, **39**, D777–D787.
 29. Yu, K., Arighi, C.N., Ross, K.E., Zhao, Q., Zhang, X., Wang, Y., Wang, Z.X., Jin, Y., Li, X., Liu, Z.X. *et al.* (2019) qPhos: a database of protein phosphorylation dynamics in humans. *Nucleic Acids Res.*, **47**, D451–D458.
 30. Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D.J., Turner, J.P., Vidovic, D., Forlin, M., Kelley, T.T., D'Urso, A. *et al.* (2017) Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res.*, **46**, D558–D566.
 31. 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.
 32. Vaudel, M., Burkhardt, J.M., Zahedi, R.P., Oveland, E., Berven, F.S., Sickmann, A., Martens, L. and Barsnes, H. (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.*, **33**, 22–24.
 33. Askenazi, M., Li, S., Singh, S. and Marto, J.A. (2010) Pathway Palette: a rich internet application for peptide-, protein- and network-oriented analysis of MS data. *Proteomics*, **10**, 1880–1885.
 34. Abelin, J.G., Patel, J., Lu, X., Feeney, C.M., Fagbami, L., Creech, A.L., Hu, R., Lam, D., Davison, D., Pino, L. *et al.* (2016) Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. *Mol. Cell Proteomics*, **15**, 1622–1641.
 35. Krug, K., Mertins, P., Zhang, B., Hornbeck, P., Raju, R., Ahmad, R., Szucs, M., Mundt, F., Forestier, D., Jane-Valbuena, J. *et al.* (2019) A curated resource for Phosphosite-specific signature analysis. *Mol. Cell Proteomics*, **18**, 576–593.