

PhosNetVis: a web-based tool for fast kinase-substrate enrichment analysis and interactive 2D/3D network visualizations of phosphoproteomics data

Osho Rawal^{1#}, Berk Turhan^{1,2#}, Irene Font Peradejordi^{1,3}, Shreya Chandrasekar^{1,3}, Selim Kalayci¹, Sacha Gnjatic^{4,5}, Jeffrey Johnson⁵, Mehdi Bouhaddou^{6,7,8}, Zeynep H. Gümüş^{*1,4}

¹ Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

² Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Türkiye

³ Cornell Tech, Cornell University, New York, NY, USA

⁴ Marc and Jennifer Lipschultz Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁵ Department of Immunology and Immunotherapy, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁶ Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁷ Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA

⁸ Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles; Los Angeles, CA, USA

⁹ Institute for Quantitative and Computational Biosciences, University of California, Los Angeles; Los Angeles, CA, USA

#co-first

*Corresponding author: zeynep.gumus@mssm.edu

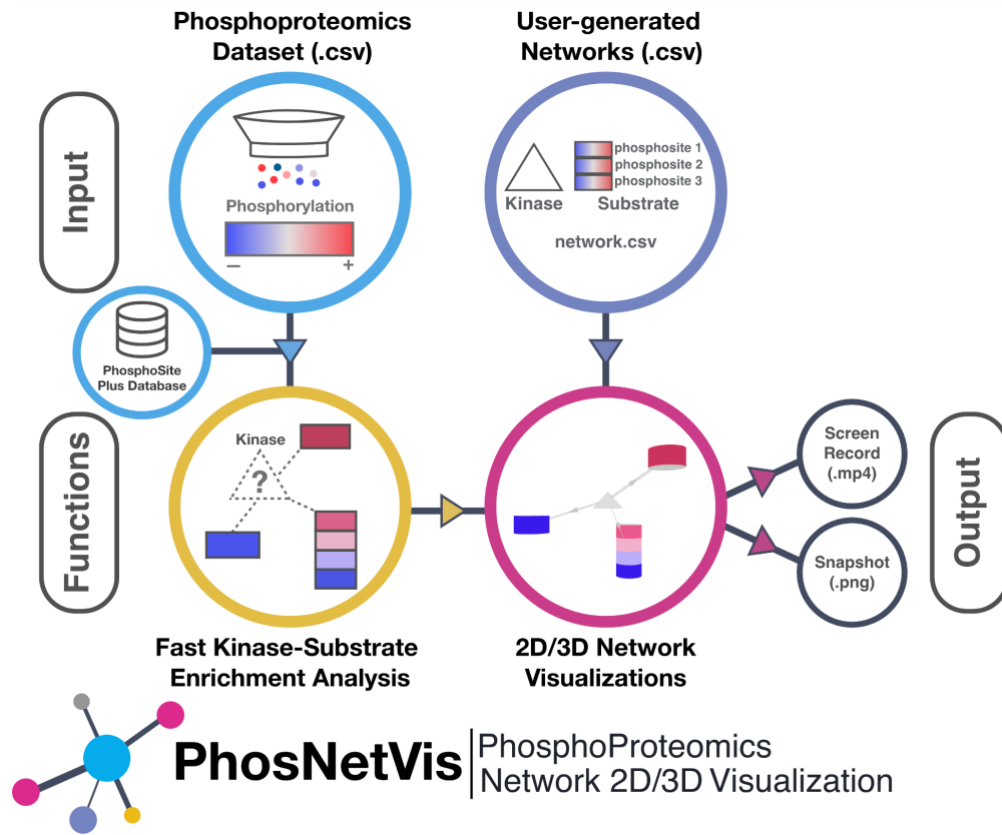
Abstract

Protein phosphorylation involves the reversible modification of a protein (substrate) residue by another protein (kinase). Liquid chromatography-mass spectrometry studies are rapidly generating massive protein phosphorylation datasets across multiple conditions. Researchers then must infer kinases responsible for changes in phosphosites of each substrate. However, tools that infer kinase-substrate interactions (KSIs) are not optimized to interactively explore the resulting large and complex networks, significant phosphosites, and states. There is thus an unmet need for a tool that facilitates user-friendly analysis, interactive exploration, visualization, and communication of phosphoproteomics datasets. We present PhosNetVis, a web-based tool for researchers of all computational skill levels to easily infer, generate and interactively explore KSI networks in 2D or 3D by streamlining phosphoproteomics data analysis steps within a single tool. PhosNetVis lowers barriers for researchers in rapidly generating high-quality visualizations to gain biological insights from their phosphoproteomics datasets. It is available at: <https://gumuslab.github.io/PhosNetVis/>

Keywords

Fast kinase-substrate enrichment analysis, kinase-substrate interaction, network visualization, interactive visualization, 3D visualization, phosphoproteomics, phosphorylation

Graphical Abstract



Introduction

Protein phosphorylation is a vital process in cellular signaling where a kinase protein modifies a residue on a substrate protein. This reversible modification can occur at multiple sites (phosphosites) on a single substrate, with different kinases targeting various substrates. Advances in liquid chromatography-mass spectrometry (LC-MS) technology have enabled the rapid generation of extensive protein phosphorylation datasets across different cellular states. To identify significant kinase-substrate interactions (KSIs) from these large datasets, bioinformatic analysis tools are essential. These tools infer which kinases are responsible for the observed changes in protein phosphorylation using available KSI database resources, and facilitate the visual exploration of the resulting KSI networks.

Over the past decade, a wide array of computational tools have been developed to infer kinases^{1,2}. Many of these tools include kinase-substrate enrichment analysis (KSEA), which employs a gene-set enrichment analysis algorithm to determine if a set of query proteins is enriched in substrates known to interact with specific kinases. Commonly employed KSEA tools are KEA3³, KSEAapp⁴, KSEA⁵, KSEAplus⁶, and pCHIPS⁷. These tools typically utilize publicly available KSI databases such as Phospho.ELM⁸, PhosphoSitePlus⁹, the Human Protein Reference Database (HPRD)¹⁰, Swissprot¹¹, and Kinase Library¹². PhosphoSitePlus⁹ is often preferred, as it is regularly updated. As of the latest update, PhosphoSitePlus lists 10,875 KSI pairs (7,895 in human), and 23,564 phosphosites (14,498 in human) (version downloaded on Thu Jun 27 2024, from the database update of Mon Jun 17 10:56:29 EDT 2024).

After performing KSEA, significant KSIs identified based on user-defined criteria are often visualized as networks, with kinases and substrates represented as nodes and their interactions as edges. Visualizing the altered phosphorylation states of phosphosites in substrate proteins is also essential. Despite the proliferation of tools to infer KSIs from phosphoproteomics datasets, currently available tools are not optimized for visualizing and generating large, complex KSI networks along with their associated phosphorylation sites and states. Furthermore, there are no dedicated tools that integrate KSEA with interactive visualizations of the resulting KSI networks and phosphosites across multiple conditions.

Current network data exploration workflows involve manual adjustments to maximize usefulness, minimize clutter, and improve visualization design. This process can be improved with scripting languages; however, proteomics researchers should not need to learn programming skills to create visualizations. Even then, tools for network visualizations typically do not include the functionalities to visually represent phosphorylation data of protein residue sites and states¹³. At the same time, tools specifically designed for visual exploration of proteomics datasets do not allow users to upload and explore their own data. For instance, ProNetView¹⁴, a web-based interactive 3D network visualization tool developed by our group, is tailored for exploring proteomics network data from the Clinical Proteogenomic Tumor Analysis Consortium (CPTAC)¹⁵. Furthermore, similar to network visualization tools, ProNetView does not visually represent the phosphorylation sites and states. One exception is the Cytoscape¹⁶ app, Omics Visualizer¹⁷, which allows visual representations of phosphosite data, but does not

include a kinase enrichment component. Finally, the popular web-based platform to investigate kinase activity data, PhosFate Profiler¹⁸, enables data exploration in tabular format only, without KSI network or phosphosite data visualization functionalities. There is thus a clear need for a tool that facilitates user-friendly generation, interactive exploration, visualization, and communication of phosphoproteomics datasets.

Here, we present PhosNetVis, a freely accessible web-based tool designed for users of all computational proficiency levels to explore kinase activity from phosphoproteomics studies. PhosNetVis integrates multiple analysis steps within a single platform, allowing users to perform KSEA and/or build and visualize shareable interactive 2D or 3D KSI networks effortlessly. The tool provides a versatile environment, offering visual representations of protein networks, phosphorylation sites, and states of interest, along with associated differential phosphorylation and statistical significance data. Users can interact with the data through KSI network visualizations and tabular formats, all within a single interface. The KSI networks can visually include phosphorylation sites and their respective states (increased or decreased). Input is straightforward, requiring users to upload their datasets as comma-separated files on their web browser. While PhosNetVis is tailored for the analysis of phosphoproteomics datasets, its adaptable architecture extends its utility to other biomolecular network applications, even in the absence of phosphorylation data. This versatility enhances its potential impact across a wide range of applications.

Results

User Research and Prototyping

Before designing and building PhosNetVis, we conducted preliminary user research with five domain practitioners. This research involved observing their current workflows, identifying pain points, and gathering requirements for ideal features. We analyzed the user input network data and how these data were typically mapped to 2D visual elements using existing tools. To establish the general workflow and convert these mappings from 2D to 3D, we first created low-fidelity (low-fi) prototypes on paper. This approach allowed us to quickly and efficiently present our assumptions on the functionalities and workflow. Following the low-fi prototypes, we developed high-fidelity (hi-fi) prototypes using Adobe XD (adobe.com/products/xd) and Figma (figma.com). We created multiple separate prototypes for both stages, which were eventually merged into a single final design. We presented the hi-fi prototype to prospective users within the Human Immunology Project Consortium (HIPC) and Clinical Proteomic Tumor Analysis Consortium (CPTAC). Based on their feedback, we made further improvements to both the prototype and the technical implementation.

PhosNetVis Workflow Design

Building KSI networks from phosphoproteomics datasets typically involves a two-step workflow: first using a KSEA tool to establish KSI networks, and then employing visualization tools to explore and analyze these networks. PhosNetVis streamlines this process by integrating both KSEA and interactive visual network exploration into a single, unified solution.

The overall architecture and main functional components of PhosNetVis are represented in Figure 1. Briefly, to infer a network of significant KSIs from a phosphoproteomics study, users input their differential phosphorylation data (log₂FC) into the fast KSEA (fKSEA) interface of PhosNetVis. For fKSEA, PhosNetVis uses the fast gene set enrichment (FGSEA) algorithm¹⁹, which allows having more permutations and thereby more fine-grained p-values than using standard multiple hypothesis correction methods. The fGSEA process generates a list of KSIs with their associated Benjamini & Hochberg (BH) corrected p-values, with a KSI considered significant if its BH-adjusted p-value is less than or equal to a user-defined cutoff. Users have the flexibility to adjust the input parameters and labels as needed.

The fKSEA page generates KSI network file(s) in .CSV format, which can then be explored in the network visualization page. On the network visualization page, users can visually explore the KSI network(s), examine phosphorylation sites and their states, query for specific proteins, adjust significance thresholds for differential phosphorylation, pan, zoom, and animate network changes across different states, such as multiple treatments or time points. Additionally, PhosNetVis allows for the direct upload of users' own phosphoproteomics network data, whether custom-built or from other KSEA tools. This feature enables users to leverage the tool's interactive visual exploration functionalities independent of its KSEA capabilities. Alternatively, users can also download the results of their fKSEA runs to visually explore the resulting network data files using other tools.

Fast Kinase-Substrate Enrichment Analysis (fKSEA) Page

A snapshot of the PhosNetVis fKSEA interface is shown in Figure 2 (descriptive text not shown). This page has three sections: input file format (Figure 2A), adjust parameters (Figure 2B) and upload data and run analysis (Figure 2C).

fKSEA Input File Format. Prior to using the fKSEA interface, users should process their proteomics data for differential phosphorylation analysis between two states (e.g., baseline vs. perturbed) to obtain log₂ fold change (perturbed/baseline) phosphorylation (log₂FC) values and associated p-values. Figure 2A illustrates the format of a sample input file, with mandatory fields marked by an asterisk (*). An input file should be in .CSV format and include at least two columns:

- Protein accession ID (identifies the protein)
- Log₂FC differential phosphorylation value at a phosphosite on that protein.

Optionally, users can include additional columns:

- Unique phosphosite ID for when multiple phosphosites are involved
- *p*-value associated with the differential phosphorylation fold change.

fKSEA Input Parameters. Users can adjust the default parameters of the FGSEA algorithm¹⁹ through the “Adjust Parameter” section (Figure 2B). These parameters include:

- Differential phosphorylation p-value cut-off (default = 0.05)
- Output BH-adjusted p-value cutoff (default=0.05)
- Minimum size of a KSI set to test (minSize) (default=15)
- Maximum size of a KSI set to test (maxSize) (default=100)
- Boundary for calculating the p-value (eps) (default=0).

For protein labels, users can choose between UniProt accession IDs or HUGO gene IDs. After uploading data and customizing parameters, the fKSEA pipeline maps the input data to KSIs within the PhosphoSitePlus database⁹ [version downloaded on Thu Jun 27 2024, from the database update of Mon Jun 17 10:56:29 EDT 2024] to identify KSIs that pass user-defined significance thresholds. Additionally, each kinase is assigned an enrichment score by the FGSEA algorithm¹⁹, which has been demonstrated in previous studies to effectively estimate kinase activities⁵.

fKSEA Output. After fKSEA runs, the resulting KSI network is output into a .CSV file. Users are then prompted with a success message that redirects them to download the KSI network connectivity data and/or to directly visualize the network in the network visualization page. A sample fKSEA output file is provided in Supplementary Table 2.

KSI Network Visualization

To visualize the KSI networks, users can either directly go from fKSEA page to the network visualization page, or skip the fKSEA step and input their datasets directly for network visualizations in the network data input page.

Network Data Input Page. For users who have generated KSIs with tools other than PhosNetVis, this page allows users to input one or more phosphoproteomics network dataset files corresponding to different cellular states (e.g. time, treatment, exposure) for KSI network visualizations. It also includes descriptive text and a formatting guide for optional customizations of the network input files. A snapshot of this page is provided in Figure 3 (descriptive text on its website not shown).

Network Data Input File Format. A network input file should include at least two columns: Kinases (Kinase ID) and Substrates (Target IDs). Users can optionally include additional attributes to customize their network visualizations according to their needs. These optional attributes include:

- KinaseSize: Node size associated with each kinase.
- KinaseActivity: Kinase node color representing positive or negative direction of kinase activity.
- EdgeWeight: Custom thickness for each edge.
- EdgeColor: Custom color for each edge.
- TargetSize: Node size associated with each substrate.
- PhosphoSiteID: Unique ID of any phosphosite users deem important to visualize.
- log2FC: Phosphosite color representing the log fold change in phosphorylation at that site.

- pValue: Associated p-value of phosphorylation.

These attributes allow users to tailor their visualizations for better clarity and insight.

KSI Network Visualization Interface. Once the KSI network files are generated in the fKSEA page, or uploaded directly through the network data input page, the user is directed to the interactive network visualization interface. PhosNetVis provides an interactive interface in optional 2D or 3D. For a comparative understanding of the 2D versus 3D layouts, Figure 4 provides two snapshots of the same KSI network generated with PhosNetVis: Panel A is a rendering in 3D layout, and Panel B shows it in 2D layout. The 2D view also includes node labels. This KSI network visualization interface features three components:

Control Panel (Figure 4, Panels A/B, left corner). This interactive panel allows users to seamlessly transition between 2D and 3D network layouts; query for specific nodes; dynamically adjust node color thresholds based on phosphorylation levels and fold-change criteria; customize the background color; toggle the visibility of phosphorylation site partitions and node labels; enter or exit fullscreen mode; capture screenshots; and restore the network visualization to default settings.

Visualized Network (Figure 4, Panels A/B middle). Users can interact with the 2D or 3D network by rotating, zooming in/out; dragging nodes to different positions or selecting nodes or edges. Substrates are represented as cylinders. The height of each cylinder depends on the number of phosphosites harbored by the substrate it represents, with each differentially phosphorylated phosphosite depicted as a slice of the cylinder. The color of each cylinder slice indicates its level of differential phosphorylation, ranging from blue (reduced phosphorylation) to red (increased phosphorylation). A maximum 10 of the most differentially altered phosphosites are shown per substrate. Kinases are represented as triangles in the 2D configuration, and as cones in the 3D configuration. For kinases, if they harbor differentially phosphorylated phosphosites, these are represented as slices below the gray triangles in the 2D configuration, colored according to the differential phosphorylation levels. In the 3D configuration, kinases without differentially phosphorylated phosphosites are shown as gray cones. If they do have such phosphosites, the cones contain slices colored according to the differential phosphorylation levels.

Pop-up node details table (Figure 4 Panel A bottom left corner). Double-clicking on a node either in 2D or 3D opens an interactive table displaying detailed information on the node. This includes the node type (kinase or substrate), and for each differentially phosphorylated site on that node, its position, differential phosphorylation fold change value (perturbation/baseline), and differential phosphorylation p-value. Additionally, it includes hyperlinks to the PhosphoSitePlus database⁹ for more detailed protein information.

Legend (Figure 4, Panels A/B, bottom right corner). This section explains the user-selected range of Phosphorylation Log Fold Change values. Each cylinder slice representing a phosphosite is colored based on the maximum negative and positive differential phosphorylation

levels in the study. Lighter hues indicate smaller values, while darker hues correspond to larger values, with blue representing the most negative and red the most positive differentially phosphorylated phosphosite. If optional parameters were included in the input file, the legend also shows the range for these attributes, such as Node Size, Edge Color, and Edge Weight (edge thickness). In the given example (Figure 4, Panels A/B), the input file includes only fold change information, so the legend displays only the range of differential fold change of phosphorylation.

This setup ensures that users can thoroughly explore and analyze their KSI networks with a high degree of interactivity and customization.

Network Visualization Output File Format. Users can take snapshots of their network visualizations, or screen-record animations of differential phosphorylation changes across multiple states. Snapshots are saved in .PNG file format. Screen recordings are saved in .MP4 file format.

Illustrative Examples

We illustrate the functionalities of PhosNetVis KSI network visualization interface with two case studies. Both are available for interactive exploration within the tool webportal. The portal also includes several toy datasets customized to help users get familiar with different KSI network attributes. By exploring these examples, users can visually interact with the networks, understand tool features and learn the correct formats for further customizing their datasets.

SARS-CoV-2 KSI network. This example is curated from a study that examined the global phosphorylation landscape of SARS-CoV-2 infection²⁰. In this study, Vero-E6 cells, an African green monkey cell line, were infected with SARS-CoV-2, the virus responsible for COVID-19. Cells were harvested at various time points post-infection (2, 4, 8, 12, and 24 hours). Phosphoproteomics analysis using liquid chromatography-mass spectrometry (LC-MS) quantified 4,624 human-orthologous phosphorylation sites across 3,036 human-orthologous proteins. A screenshot of this KSI network at the 24 hour time point is shown in Figure 4, with Panel A in 3D and Panel B in 2D.

One significant finding in this study was the increased activity of Casein Kinase II (*CSNK2A1*) following infection. This was evidenced by a marked increase in the abundance of known *CSNK2A1* phosphorylation sites post-infection. *CSNK2A1* was localized within filopodia protrusions that emerged from the cell surface during infection, suggesting a role in filopodia formation. Further analysis of phosphorylation sites affecting cytoskeletal changes revealed increased phosphorylation of *CTNNA1*, *HDAC2*, *HMGA1*, *HMGN1*, and *STAT1* proteins at sites known to be associated with cytoskeleton remodeling. Specifically, phosphorylation was observed at *CTNNA1:S641*, *HDAC2:S394*, *HMGA1:S102-103*, *HMGN1:S7*, *STAT1:S727*. Zooming in on *CSNK2A1* in this network and inspecting its substrates and known phosphosites visually confirms this finding. These interactions were identified by running the fKSEA algorithm on the phosphoproteomics data, and are depicted in the magnified subnetwork of *CSNK2A1* (Figure 4, Panels A/B, zoomed-in circles). Clicking on one of its substrates, *HMGA1*, brings up a

pop-up table of its differentially phosphorylated phosphosites, S102 and S103, along with their log₂ fold change (infection/mock) and p-values (Figure 4, Panel A, bottom left). For further details on *HMGA1*, the table links to PhosphoSitePlus⁹.

This KSI network is available for interactive exploration in PhosNetVis at:

<https://gumuslab.github.io/PhosNetVis/existing-networks.html>. Users can inspect the phosphorylation status of the network at a time point of interest by simply using the dropdown menu in the Control Panel (Figure 4, Panels A/B, top left), toggle across different time points, or inspect an animation of the changes in phosphorylation within the KSI network over time by selecting the *Animate & Record* tab from the Control Panel. These functionalities help users pinpoint potential mechanisms of interest for further studies, providing a dynamic and detailed view of the KSI network and the phosphorylation events during SARS-CoV-2 infection.

Clinical Proteomic Tumor Analysis Consortium (CPTAC) pan-cancer immune subtype KSI networks. Recently, through the Clinical Proteomic Tumor Analysis Consortium (CPTAC), we investigated the immune landscape of over 1,000 tumors from ten different cancer types²¹. This effort aimed to enhance the understanding of immune cell surveillance mechanisms and the various strategies tumors use to evade immune responses. Using CPTAC's comprehensive pan-cancer proteogenomic data²², this study identified and characterized seven unique immune subtypes, and by analyzing kinase activities within these subtypes, it uncovered potential therapeutic targets specific to each subtype. Here, we provide the full catalog of PhosNetVis interactive visualizations of KSI networks of each immune subtype both as a resource for the research community and as use case examples. This PhosNetVis application skipped the fKSEA Page. Instead, the networks were derived from simply mapping the list of proteins (both kinases and substrates) in each immune subtype to the PhosphoSitePlus database⁹ to derive their connecting edges between the kinases and the phosphosites in the CPTAC immune study²¹. Then, the adjacency matrices of all immune subtype KSI networks were directly uploaded into PhosNetVis using its Network Data Input Page to generate their interactive visualizations.

The PhosNetVis catalog of these networks allows easy queries, interactive visualizations, exploration, and download, and is available through the PhosNetVis Existing Networks link at: <https://gumuslab.github.io/PhosNetVis/cptac-vis.html>. Using PhosNetVis, users can on-the-fly comparatively analyze KSIs and differential phosphorylations across these 7 immune subtypes. For example, Figure 5 illustrates two of the largest-connected network snapshots in a 2D layout. *Panel A* shows the most active pan-cancer immune subtype (CD8+/IFNG+), which leads to the upregulation of PRKACA and the phosphorylation of its substrates. *Panel B* depicts the least active cluster (CD8-/IFNG-), resulting in the upregulation of cell-cycle kinases such as CDK1 and the phosphorylation of its substrates.

These CPTAC networks provide good case studies on how users can interact with the networks in 2D or 3D layout, and then directly embed their network snapshots in 2D into publications, as we illustrate in Figure 5. The drag function allows users to easily modify the network layout, emphasizing specific parts of the network as needed. In Figure 5, Panels A and B, quick manual alterations in the network connectivities and layouts highlight the main kinases more clearly.

These visualizations enable researchers to pinpoint potential mechanisms of interest for follow-up studies. Furthermore, the catalog of these pan-cancer immune subtype KSI networks enables the broader community of researchers to explore complex KSI network datasets from the CPTAC initiative.

Discussion

Large-scale phosphoproteomics experiments are characterizing protein phosphorylation sites and states across conditions to better understand cellular signaling in health and disease. While numerous tools are dedicated to infer kinases from these datasets, there is a need for integrated visualization tools that allow users to explore the resulting KSI networks across multiple conditions. Here, we introduced PhosNetVis, an interactive web-based tool that enables users to perform fast Kinase-Substrate Enrichment Analysis (fkSEA), and then to visually explore, interpret and communicate the inferred KSI networks and their associated phosphorylation data across multiple states. PhosNetVis provides a private analysis environment, as all operations during KSI network visualization take place exclusively within the user's local browser, without relying on an external server or third-party site. In the case of fkSEA, datasets undergo processing on the protected RStudio Connect server at Mount Sinai (<https://rstudio-connect.hpc.mssm.edu>) to guarantee data privacy. By integrating fkSEA and network visualization in a single platform, PhosNetVis simplifies the workflow for analyzing phosphoproteomics data, making it accessible to users of varying computational proficiency levels.

PhosNetVis features an easy-to-navigate graphical user interface for KSI networks that visually represents phosphorylation sites and their respective states across multiple experiments. Users can explore KSI network visualizations, switch between different network states (i.e. time points or treatments), toggle between 2D or 3D visualizations, and adjust network parameters. The tool allows users to query nodes of interest for detailed phosphosite information, change differential phosphorylation parameters, rotate, pan, zoom in/out, drag nodes, download the network snapshots or record animations of network changes over time or conditions. While PhosNetVis is specifically designed for phosphoproteomics analysis, its adaptable architecture extends its utility to various biomolecular network applications, even when phosphorylation data are unavailable. This adaptability broadens its potential impact across diverse research problems, enabling researchers to visualize and interpret complex biological data in a dynamic and interactive manner. One limitation is that while fkSEA analysis employs the most recent version of the most popularly used kinase-substrate database, PhosPhoSitePlus⁹, the upstream kinases of a majority of phosphosites are still unknown. Furthermore, 20% of those kinases that are known are in the upstream of the majority of known phosphosites²³. This, however, is a limitation within the field, and is an active area of research, with studies ongoing to reveal the full extent of human KSIs.

Overall, PhosNetVis lowers the barriers between complex phosphoproteomics data and researchers who want rapid, intuitive, and high-quality tools to infer kinase activity and thereby visually explore kinase-substrate interaction networks at multiple phosphorylation sites and

states. This will empower investigators in translating rich datasets into biological insights and clinical applications. PhosNetVis is freely accessible <https://gumuslab.github.io/PhosNetVis>.

Experimental Procedures

Resource availability

Lead contact

Dr. Zeynep H. Gümüş can be reached by email (zeynep.gumus@mssm.edu).

Materials availability

PhosNetVis is hosted on GitHub Pages and is freely available with no login requirements at <https://gumuslab.github.io/PhosNetVis>. For additional materials, please contact the lead contact.

Data and code availability

PhosNetVis source code is available on GitHub repository (github.com/GumusLab/PhosNetVis), and is protected with Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License CC-BY-NC-SA-4.0 (creativecommons.org/licenses/by-nc-sa/4.0/legalcode).

Website Implementation

fKSEA Page. We developed a user-friendly interface for fKSEA using HTML, CSS, and Bootstrap 4. FGSEA is performed using a Plumber API (<https://www.rplumber.io/>), deployed on the RStudio Connect server at Mount Sinai (<https://rstudio-connect.hpc.mssm.edu>). Users upload their CSV files and initiate the analysis via a simple interface, which sends an API request to the server to perform FGSEA. This setup ensures efficient and seamless processing of kinase enrichment analysis, making it accessible and convenient for researchers. Once fKSEA API call is finished, users are directed to download the network data files or to directly visualize the files..

Network Data Input Page. Users can upload one or more network data files through this page. Network data files can be the outputs from PhosNetVis fKSEA page analysis or custom-generated. The page provides detailed guidelines for network data input formats for custom-generated files. For CSV file parsing and data processing, we integrated the PapaParse library (<https://www.papaparse.com/>) to enable users to concurrently upload and process multiple CSV files. Once network data CSV files are uploaded, PhosNetVis transforms each input .CSV file into the JavaScript Object Notation (JSON) format using the Danfo.JS package (danfo.jsdata.org/).

Interactive Network Visualization Page. Once network data files are uploaded, all networks are visualized in the network visualization page. We built this interface mainly with HTML and CSS. We used JavaScript for the Document Object Model (DOM) element manipulations, and the Bootstrap library to implement responsive layouts and to customize page architectures (getbootstrap.com). To create a screen-resizable responsive canvas where the 3D network

visualization is rendered, we integrated the element-resize-detector library (github.com/wnr/element-resize-detector)²⁴.

To prioritize user interactions and control, we included several graphical user interface (GUI) elements to the network visualization interface (Figure 4, Panels A and B). Briefly, to enable users to interactively modify their visualizations, adjust parameters and conduct node queries, we implemented a GUI control panel by using the Tweakpane.js library (cocopon.github.io/tweakpane/) (Upper left corner in Figure 4, Panels A and B). To enhance user interaction through specific node queries, we used the Fstdropdown.js library (github.com/VirtusX/fstdropdown). To offer users a 3D network visualization experience (Figure 4, Panel A), we used the 3d-force-graph library (github.com/vasturiano/3d-force-graph) to render the network layouts in 3D. The library leverages the WebGL-based Three.js (threejs.org) library to create interactive force-directed graphs in 3D space. Users can easily customize node colors and labels in 3D, for which we used RainbowVis-JS (github.com/anomal/RainbowVis-JS) and three-spritetext (github.com/vasturiano/three-spritetext) libraries, respectively. In addition to 3D, PhosNetVis also offers a 2D network visualization option (Figure 4, Panel B). To develop the 2D network visualization option, we implemented the force-graph library (github.com/vasturiano/force-graph), which uses a force-directed layout algorithm to present network layouts effectively on an HTML5 canvas. Finally, for users to export visualizations, we implemented jsScreenRecorder (github.com/manan657/jsScreenRecorder), a custom JavaScript script designed to capture and record screen interactions, particularly beneficial after rendering the networks in 3D. By implementing these tools and libraries, PhosNetVis ensures a comprehensive and user-friendly network visualization experience, supporting researchers in their exploration of KSI networks and associated phosphorylation data.

Tutorials, Interactive Examples and Toy Datasets

To guide users, the documentation platform of PhosNetVis (<https://gumuslab.github.io/phosnetvis-docs/>) features tutorials that cover fKSEA and network visualization, as well as instructions on input data formatting. In addition, we provide several interactive examples and toy datasets at github.com/GumusLab/PhosNetVis-DataExamples.

Acknowledgments

ZHG gratefully acknowledges support from NIH R33 CA263705-01. Authors gratefully acknowledge valuable feedback from investigators within NIH funded Human Immunology Project Consortium (HIPC) and Clinical Proteomic Tumor Analysis Consortium (CPTAC), with special thanks to Francesca Petralia for kindly sharing the CPTAC pan-cancer immune subtype networks with the study team. MB acknowledges support from NIH K99 AI163868.

Author contributions

O.R: writing - reviewing & editing, implementation, visualization, software, investigation. B.T: writing - original draft, writing - reviewing & editing, implementation, visualization, software, investigation. I.F.P: writing - original draft, implementation, visualization, software, investigation.

S.C: writing - original draft, implementation, visualization, software, investigation. S.G: writing - reviewing, visualization. S.K: implementation, visualization, software. J. J: writing - original draft, writing - reviewing & editing, visualization, software, conceptualization. M.B.L: writing- reviewing, visualization. Z.H.G: writing - original draft, writing - reviewing & editing, conceptualization, visualization, supervision.

Declaration of interests

The authors declare no competing interests.

References

1. Hallal, M., Braga-Lagache, S., Jankovic, J., Simillion, C., Bruggmann, R., Uldry, A.-C., Allam, R., Heller, M., and Bonadies, N. (2021). Inference of kinase-signaling networks in human myeloid cell line models by Phosphoproteomics using kinase activity enrichment analysis (KAEA). *BMC Cancer* 21, 789. <https://doi.org/10.1186/s12885-021-08479-z>.
2. Piersma, S.R., Valles-Marti, A., Rolfs, F., Pham, T. V., Henneman, A.A., and Jiménez, C.R. (2022). Inferring kinase activity from phosphoproteomic data: Tool comparison and recent applications. *Mass Spectrom Rev*, 725–751. <https://doi.org/10.1002/mas.21808>.
3. Kuleshov, M. V, Xie, Z., London, A.B.K., Yang, J., Evangelista, J.E., Lachmann, A., Shu, I., Torre, D., and Ma'ayan, A. (2021). KEA3: improved kinase enrichment analysis via data integration. *Nucleic Acids Res* 49, W304–W316. <https://doi.org/10.1093/nar/gkab359>.
4. Wiredja, D.D., Koyutürk, M., and Chance, M.R. (2017). The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* 33, 3489–3491. <https://doi.org/10.1093/bioinformatics/btx415>.
5. Hernandez-Armenta, C., Ochoa, D., Gonçalves, E., Saez-Rodriguez, J., and Beltrao, P. (2017). Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics* 33, 1845–1851. <https://doi.org/10.1093/bioinformatics/btx082>.
6. Casado, P., Hijazi, M., Gerdes, H., and Cutillas, P.R. (2022). Implementation of Clinical Phosphoproteomics and Proteomics for Personalized Medicine. In, pp. 87–106. https://doi.org/10.1007/978-1-0716-1936-0_8.
7. Drake, J.M., Paull, E.O., Graham, N.A., Lee, J.K., Smith, B.A., Titz, B., Stoyanova, T., Faltermeier, C.M., Uzunangelov, V., Carlin, D.E., et al. (2016). Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell* 166, 1041–1054. <https://doi.org/10.1016/j.cell.2016.07.007>.
8. 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. <https://doi.org/10.1093/nar/gkq1104>.
9. 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. <https://doi.org/10.1093/nar/gku1267>.
10. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009).

- Human Protein Reference Database--2009 update. *Nucleic Acids Res* 37, D767–D772. <https://doi.org/10.1093/nar/gkn892>.
11. Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., et al. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49, D480–D489. <https://doi.org/10.1093/nar/gkaa1100>.
 12. Johnson, J.L., Yaron, T.M., Huntsman, E.M., Kerelsky, A., Song, J., Regev, A., Lin, T.-Y., Liberatore, K., Cizin, D.M., Cohen, B.M., et al. (2023). An atlas of substrate specificities for the human serine/threonine kinome. *Nature* 613, 759–766. <https://doi.org/10.1038/s41586-022-05575-3>.
 13. Liliashvili, V., Kalayci, S., Fluder, E., Wilson, M., Gabow, A., and Gümüş, Z.H. (2017). iCAVE: an open source tool for visualizing biomolecular networks in 3D, stereoscopic 3D and immersive 3D. *Gigascience* 6. <https://doi.org/10.1093/gigascience/gix054>.
 14. Kalayci, S., Petralia, F., Wang, P., and Gümüş, Z.H. (2020). ProNetView-ccRCC: A Web-Based Portal to Interactively Explore Clear Cell Renal Cell Carcinoma Proteogenomics Networks. *Proteomics* 20. <https://doi.org/10.1002/pmic.202000043>.
 15. Petralia, F., Tignor, N., Reva, B., Koptyra, M., Chowdhury, S., Rykunov, D., Krek, A., Ma, W., Zhu, Y., Ji, J., et al. (2020). Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* 183, 1962-1985.e31. <https://doi.org/10.1016/j.cell.2020.10.044>.
 16. 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. <https://doi.org/10.1101/gr.1239303>.
 17. Legeay, M., Doncheva, N.T., Morris, J.H., and Jensen, L.J. (2020). Visualize omics data on networks with Omics Visualizer, a Cytoscape App. *F1000Res* 9, 157. <https://doi.org/10.12688/f1000research.22280.2>.
 18. Ochoa, D., Jonikas, M., Lawrence, R.T., El Debs, B., Selkrig, J., Typas, A., Villén, J., Santos, S.D., and Beltrao, P. (2016). An atlas of human kinase regulation. *Mol Syst Biol* 12. <https://doi.org/10.15252/msb.20167295>.
 19. Gennady Korotkevich, V.S.N.B.B.S.M.N.A.A.S. (2021). Fast gene set enrichment analysis. Preprint, <https://doi.org/https://doi.org/10.1101/060012> <https://doi.org/https://doi.org/10.1101/060012>.
 20. Bouhaddou, M., Memon, D., Meyer, B., White, K.M., Rezelj, V. V., Correa Marrero, M., Polacco, B.J., Melnyk, J.E., Ulferts, S., Kaake, R.M., et al. (2020). The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* 182, 685-712.e19. <https://doi.org/10.1016/j.cell.2020.06.034>.
 21. Petralia, F., Ma, W., Yaron, T.M., Caruso, F.P., Tignor, N., Wang, J.M., Charytonowicz, D., Johnson, J.L., Huntsman, E.M., Marino, G.B., et al. (2024). Pan-cancer proteogenomics characterization of tumor immunity. *Cell* 187, 1255-1277.e27. <https://doi.org/10.1016/j.cell.2024.01.027>.
 22. Li, Y., Dou, Y., Da Veiga Leprevost, F., Geffen, Y., Calinawan, A.P., Aguet, F., Akiyama, Y., Anand, S., Birger, C., Cao, S., et al. (2023). Proteogenomic data and resources for

- pan-cancer analysis. *Cancer Cell* 41, 1397–1406.
<https://doi.org/10.1016/j.ccell.2023.06.009>.
23. Needham, E.J., Parker, B.L., Burykin, T., James, D.E., and Humphrey, S.J. (2019). Illuminating the dark phosphoproteome. *Sci Signal* 12.
<https://doi.org/10.1126/scisignal.aau8645>.
24. Lucas Wiener, T.E.P.H. (2015). Modular Responsive Web Design using Element Queries. Preprint, <https://doi.org/10.48550/arXiv.1511.01223>.

Figures

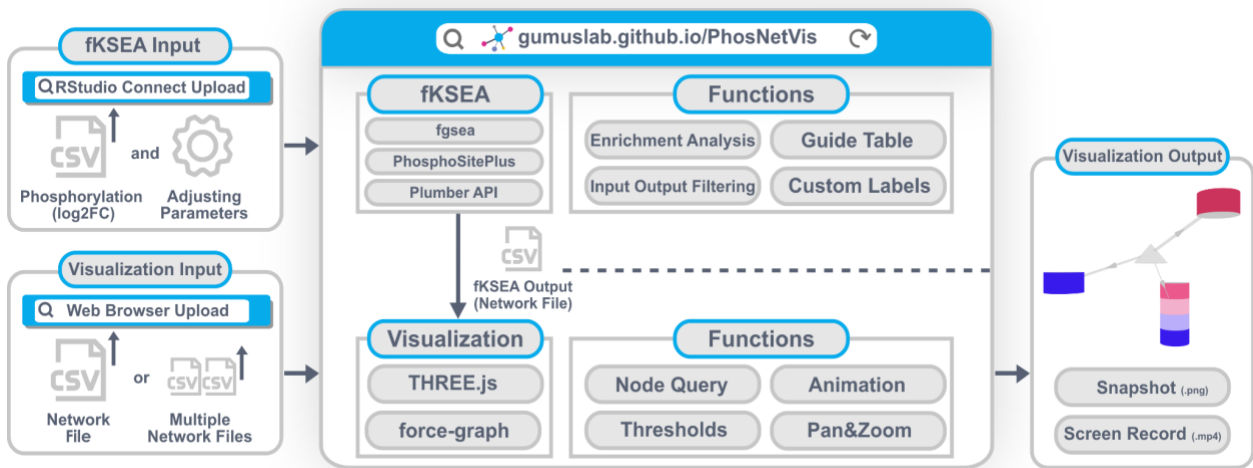


Figure 1. PhosNetVis overall architecture and main components.

A Input File Format

Upload a .csv file in the format below for enrichment and then to build your interactive 3D phosphoproteomics network! *optional

ProteinAccession	PhosphoSiteID*	log2FC	pValue*
P54646	ph181	1.2	0.002
P54646	ph14	3	0.04
Q9Y243	ph1525	-6.55	0.66

B Adjust Parameters

Input / Output Parameters

Input p-value cutoff:

Output p-value cutoff:

Output network node labels: gene name accession id

Enrichment Analysis Parameters

minSize:

maxSize:

eps:

[Click here for more information about fgsea library's analysis parameters](#)

C Upload Data & Run Analysis

Upload your .csv file

Choose File kea_tutorial.csv

Upload Complete

Figure 2. fKSEA page snapshot. A.) Input File Format section describes input .CSV file format for fKSEA input. B.) Adjust Parameters section enables users to adjust parameters based on their needs before running the FGSEA algorithm. C.) Upload Data & Run analysis section enables users to upload their input file from their local directory; run fKSEA; download the generated KSI network file and visualize the generated network.

KinaseID (Required)	KinaseSize (Optional)	KinaseActivity (Optional)	EdgeWeight (Optional)	EdgeColor (Optional)	TargetID (Required)	TargetSize (Optional)	PhosphoSiteID (Optional)	log2FC (Optional) (Phosphorylation)	pValue (Optional) (Phosphorylation)
CDK1	1.2	0.16	1	1.44	NUCKS1	0.9	S19	-0.1	0.05
CDK1	1.2	0.16	1	1.44	NUCKS1	0.9	T202	-4.32	0.006
CDK1	1.2	0.16	1.5	4.28	CDK1	1.2	T161	1.47	0.12
CDK1	3	0.16	3.8	0.07	NPM1	1.12	T78	2	0.0235
CDK1	3	0.16	3.8	0.07	NPM1	1.12	S106	1.75	0.1
CSNK2A1	4.56	-2.27	1.67	2	TOP2A	2.98	S29	-3.059	0.03
CSNK2A1	4.56	-2.27	1.67	2	TOP2A	2.98	S285	1.434	0.27
MAPK3	1	1.2	2.28	1.5	TOP2A	2.98	K431	-2.51	0.01
MAPK3	1	1.2	0.3	7.54	RRAS2	1.77	T602	2.177	0.074

If you have more than one dataset, select the number of datasets and make sure to upload all of them!

Number of Datasets: No file chosen

UPLOAD

or

**FAST KINASE-SUBSTRATE
ENRICHMENT ANALYSIS**



TUTORIAL

Figure 3. Snapshot of the Network Visualization Input Page (descriptive text not shown). This page allows users to upload network files for visualization. Users can upload their custom network files or those generated by the fKSEA page. Users can upload multiple files by selecting the number of datasets they want to upload. The page also displays a table that provides the required network data format. To visualize network data, users need a .csv file with at least two columns: 1) Kinases (KinaseID) and 2) Target Nodes (TargetID) for substrates. Optionally, users can add additional columns to customize kinase nodes, target nodes and edge attributes.

A PhosNetVis

Controls Animate & Record

Network: time 24hr

Pos-Log2FC:

Neg-Log2FC:

Background: #ffffff

Partitions: Hide Show

Labels: On Off

Enter/Exit Fullscreen

Take a screenshot

Reset

Switch to 2D

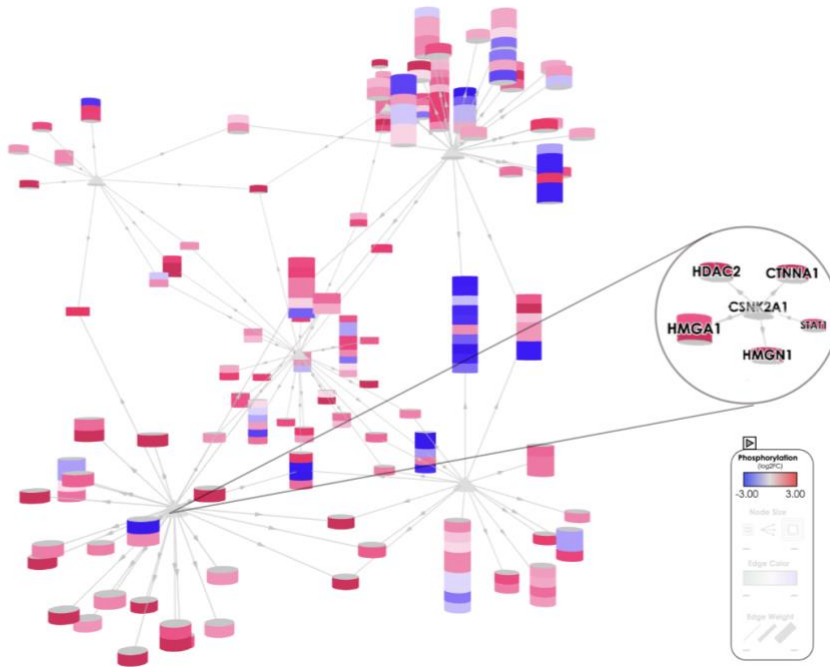
Node Query

Node: HMG1

Type: Substrate

Phosphorylation Info *(scrollable)

Site	log2FC	pValue
S102	2.050	0.001
S103	3.190	<0.001



B PhosNetVis

Controls Record

Network: time 24hr

Pos-Log2FC:

Neg-Log2FC:

Background: #ffffff

Partitions: Hide Show

Labels: On Off

Enter/Exit Fullscreen

Take a screenshot

Reset

Switch to 3D

Node Query

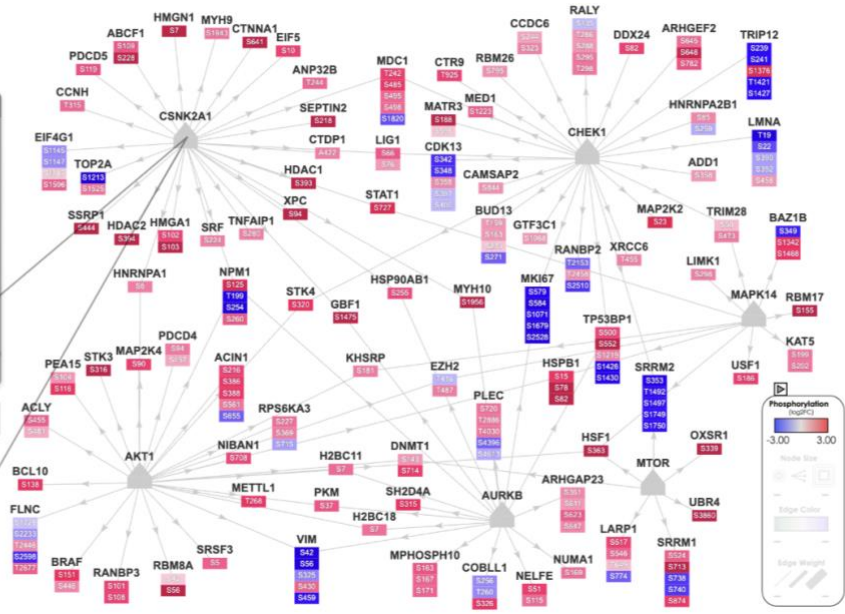
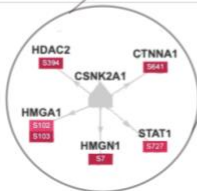


Figure 4. Snapshots of network visualization interface featuring the phosphorylation landscape of SARS-CoV-2 infection. This page allows users to view and interact with the network curated from the Global Phosphorylation Landscape of SARS-CoV-2 Infection (Bouhaddou et al., 2020). Users can rotate the network, zoom in/out, pan through, see a node information by double clicking on it, or reset the view by double clicking on the background. Additionally, it provides options for the user to switch between 3D or 2D network layouts.

Double-clicking on any node opens an interactive table that displays detailed information on the node as shown in Panel A bottom left for node *HMGGA1*. **A)** Snapshot of the 3D view of the network, including a magnified view of a CSNK2A1-central subnetwork in 3D. This subnetwork highlights phosphorylation sites known to be associated with cytoskeleton remodeling, filtered from the main network. **B)** Snapshot of the 2D view of the same network. In each panel, the left corner shows the control panel; the middle displays the network; the bottom right corner contains the legend. The starred section indicates the CSNK2A1 phosphorylation subnetwork, including a magnified view of the same CSNK2A1 node with select interactors in a subnetwork in 2D. These features enable comprehensive exploration and analysis of the phosphorylation events during SARS-CoV-2 infection.

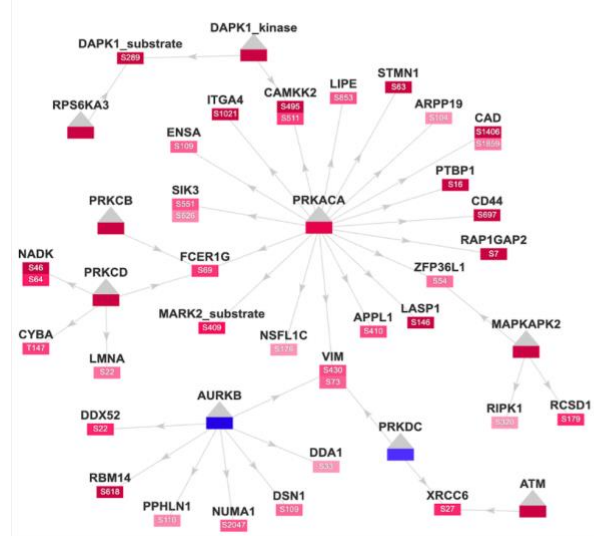
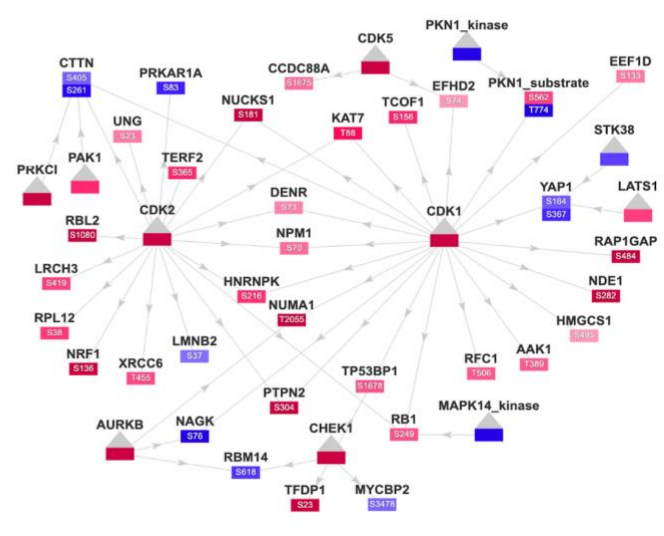
A) CD8+/IFNG+**B) CD8-/IFNG-**

Figure 5. Snapshots of interactive 2D network visualizations for CPTAC pan-cancer immune subtypes. Kinases, and their specific altered phosphosites, their phosphorylation levels and positions in each altered substrate are clearly shown **A) CD8+/IFNG+ network.** Visualization clearly depicts upregulation in the global abundance of the kinase PRKACA (center) and the phospho-abundance of its substrates in red; **B) CD8-/IFNG- subnetwork.** The global proteomic expression of cell-cycle kinases such as CDK1 and the phospho-abundance of its substrates are upregulated in red.

SUPPLEMENTARY TABLES

Supplementary Table 1. Fast Kinase-Substrate Enrichment Analysis (fKSEA) Input File Format. [(R): Required, (O): Optional]

ProteinAccession (R)	PhosphoSiteID (O)	log2FC (R)	pValue (O)
P02545	S22	-1.8	0.0025
P02545	S390	-0.95	0.03
P17612	T198	2.1	0.028
P08670	S56	-3.48	0.024

Supplementary Table 2. Fast Kinase-Substrate Enrichment Analysis (fKSEA) Output File Format. [(H|A): HUGO or Accession ID, (IE): If Exists]

KinaseID (H A)	TargetID (H A)	PhosphoSiteID (IE)	log2FC	KinaseActivity	pValue
AKT1	LMNA	S22	-1.8	-1.3	0.0025
AKT1	LMNA	S390	-0.95	-1.3	0.03
PRKACA	PRKACA	T198	2.1	0.7	0.028
PRKACA	VIM	S56	-3.48	0.7	0.024

Supplementary Table 3. Visualization Input File Format. [(R): Required, (O): Optional, (P): Phosphorylation]

Kinase ID (R)	Kinase Size (O)	Kinase Activity (O)	Edge Weight (O)	Edge Color (O)	Target ID (R)	Target Size (O)	PhosphoSite ID (O)	log2FC (O) (P)	pValue (O) (P)
AKT1	1.2	-1.3	1	1.44	LMNA	0.9	S22	-1.8	0.0025
AKT1	1.2	-1.3	1	1.44	LMNA	0.9	S390	-0.95	0.03
PRKACA	3	0.7	1.5	4.28	PRKACA	3	T198	2.1	0.028
PRKACA	3	0.7	3.8	0.07	VIM	1.12	S56	-3.48	0.024