

<https://doi.org/10.1038/s42003-025-07548-4>

A computational tool to infer enzyme activity using post-translational modification profiling data



Dehui Kong^{1,2}, Aijun Zhang^{1,2}, Ling Li^{1,2}, Zuo-Fei Yuan³, Yingxue Fu^{1,2}, Long Wu³, Ashutosh Mishra³, Anthony A. High³, Junmin Peng^{4,5}✉ & Xusheng Wang^{1,2}✉

Enzymes play a pivotal role in orchestrating complex cellular responses to external stimuli and environmental changes through signal transduction pathways. Despite their crucial roles, measuring enzyme activities is typically indirect and performed on a smaller scale, unlike protein abundance measured by high-throughput proteomics. Moreover, it is challenging to derive the activity of enzymes from proteome-wide post-translational modification (PTM) profiling data. To address this challenge, we introduce enzyme activity inference with structural equation modeling under the JUMP umbrella (JUMPsem), a novel computational tool designed to infer enzyme activity using PTM profiling data. We demonstrate that the JUMPsem program enables estimating kinase activities using phosphoproteome data, ubiquitin E3 ligase activities from the ubiquitinome, and histone acetyltransferase (HAT) activities based on the acetylome. In addition, JUMPsem is capable of establishing novel enzyme-substrate relationships through searching motif sequences. JUMPsem outperforms widely used kinase activity tools, such as IKAP and KSEA, in terms of the number of kinases and the computational speed. The JUMPsem program is scalable and publicly available as an open-source R package and user-friendly web-based R/Shiny app. Collectively, JUMPsem provides an improved tool for inferring protein enzyme activities, potentially facilitating targeted drug development.

Protein enzymes are crucial in regulating cellular function, orchestrating activities through intricate signal transduction pathways¹. Dysregulation in enzyme activity has been implicated in numerous complex human diseases, such as cancer, diabetes, and neurodegeneration^{2–4}. Given their central role in disease pathways, several groups of enzymes, including protein kinases, have emerged as the primary targets for cancer drugs^{5–7}. However, responses to the drugs can vary significantly among individual patients, underscoring the necessity for accurate measurement of enzyme activities. Traditionally, enzyme activities have been primarily characterized through substrate events and assessed via small-scale in vitro experiments^{8,9}. With advancements in high-throughput mass spectrometry (MS)-based proteomics, large-scale proteome profiling has become a powerful tool for measuring post-translational modification (PTM) events and understanding signal transduction pathways in diseases^{10,11}. For instance, these studies identified 19,947 relationships of kinase-phosphorylated substrates in human, mouse,

and rat¹², 3556 relationships of ligase-ubiquitinated substrates in human, mouse, and rat¹³, and 503 human relationships of histone acetyltransferases (HAT)-acetylated substrate¹⁴ on a large scale. Despite these advancements, the analysis of PTM profiling data remains challenging due to the lack of effective tools for inferring the activities of enzymes, such as protein kinases, E3 ubiquitin ligases, and HATs.

To address the challenges, several tools have been developed to derive activities for protein kinases using phosphoproteomics data. However, no tools are currently available for E3 ubiquitin ligases and HATs. These tools can be broadly categorized into three main types based on their underlying algorithms: enrichment-based methods, clustering algorithms, and machine-learning methods. Enrichment-based methods, including GSEA¹⁵, KSEA¹⁶, KEA¹⁷, and VIPER¹⁸, predict kinase activity by leveraging kinase-substrate interactions obtained from public databases (e.g., PhosphoSite-Plus). While a major limitation of these methods is their reliance on prior

¹Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA. ²Department of Neurology, University of Tennessee Health Science Center, Memphis, TN, USA. ³Center for Proteomics and Metabolomics, St. Jude Children's Research Hospital, Memphis, TN, USA.

⁴Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA. ⁵Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN, USA. ✉e-mail: junmin.peng@stjude.org; xwang39@uthsc.edu

knowledge of kinase-substrate relationships in databases, some extensions have been developed that use unbiased, fully data-driven approaches. These include the reverse-engineering of regulatory models from experimental proteomics data¹⁹ and even from expression profiles using ARACNe and VIPER^{18,20}. However, a major limitation of these methods is their reliance on prior knowledge of kinase-substrate relationships in databases. Clustering algorithms (e.g., CLUE) group phosphoproteomics data into clusters based on similar spatio-temporal kinetics of the phosphosites followed by an enrichment test. However, these methods often incorporate potential ‘noise’, such as low-abundance phosphopeptides and weak correlation between kinase-substrate relationships. Machine-learning methods (e.g., IKAP) attempt to globally fit a model of kinase activity and affinity to the phosphoproteomics data. However, these methods may face challenges in achieving a global optimum solution, and the derived activities from the model can vary with each computation.

Structural equation modeling (SEM) has proven to be a powerful method and is extensively employed for analyzing biological networks in genomics²¹. However, its application in proteomics, particularly for enzyme activity inference, remains largely unexplored. In contrast to existing tools for inferring protein enzyme activities, SEM offers several advantages. Firstly, it enables the inference of latent variables, such as protein enzyme activities, which cannot be directly measured. Secondly, SEM accounts for the interactions among enzymes, aligning closely with the inherent complexity of biological systems. Lastly, it incorporates measurement errors in all observed variables, such as an abundance of phosphopeptides in phosphoproteome experiments. This is especially true in MS-based PTM profiling data, where measurements inherently include errors and imperfections arising from biological and experimental variations.

In this study, we aim to develop a computational tool of JUMPsem for inferring the activities of protein enzymes. The program infers enzyme activities based on enzyme-substrate relations and PTM profiling data. We then apply the JUMPsem tool to analyze kinase activities using phosphoproteome data, E3 ligase activities using the ubiquitinome, and HAT activities using the acetylome. To assess its performance, we compare JUMPsem with two widely used programs, IKAP and KSEA, for protein kinase activity. Another feature of the JUMPsem program is its ability to incorporate sequence motifs, significantly enhancing the identification of enzyme-substrate relationships. The JUMPsem program is made publicly

available as an open-source R package, along with a user-friendly web-based R/Shiny application.

Results

Design and implementation of the JUMPsem program

We developed JUMPsem, a software tool specifically designed to infer protein enzyme activity using MS-based PTM profiling data (Fig. 1). The JUMPsem program comprises three major components: (i) construction of enzyme-substrate relationships, (ii) inference of enzyme activity through a structural equation modeling algorithm, and (iii) output and comparison of enzyme activity. The program takes a quantitative table of modified peptides from PTM profiling studies as input, which can include analyses of phosphoproteome, ubiquitinome, or acetylome. Outputs from JUMPsem consist of a table containing inferred enzyme activities and a second table containing the affinities between protein enzymes and their substrates (Supplementary Fig. S1).

The quantitative table of PTM peptides can be generated by identifying peptides via database search, filtering PTM peptides at a user-specified false discovery rate (FDR), and quantifying PTM peptides from either labeled or label-free PTM profiling data. If the corresponding whole proteome data is available, JUMPsem provides an option for normalizing the measured expression levels of PTM peptides relative to the corresponding changes in protein abundance. This normalization allows the determination of changes in PTM state independent of variations at the protein level (Supplementary Fig. S2). In the case of phosphoproteome analysis, the JUMPsem program initiates the process by extracting kinase-substrate relationship tables from public databases, such as the PhosphoSitePlus (PSP) database¹². Subsequently, the program constructs an adjacency matrix for those kinase-substrate relationships with detected phosphopeptides identified in the quantitative phosphoproteomics data. The adjacency matrix can be expanded to include kinase-substrate relationships derived from motif searches, which increases the depth of analyzed kinases (Supplementary Fig. S3).

The JUMPsem program uses the SEM algorithm to calculate enzyme activity, implemented through the lavaan R package²². The core of its computation lies in the “sem” function, which fits the model by applying maximum likelihood estimation to determine the parameters. Both the adjacency matrix and PTM abundance table are used for the estimation.

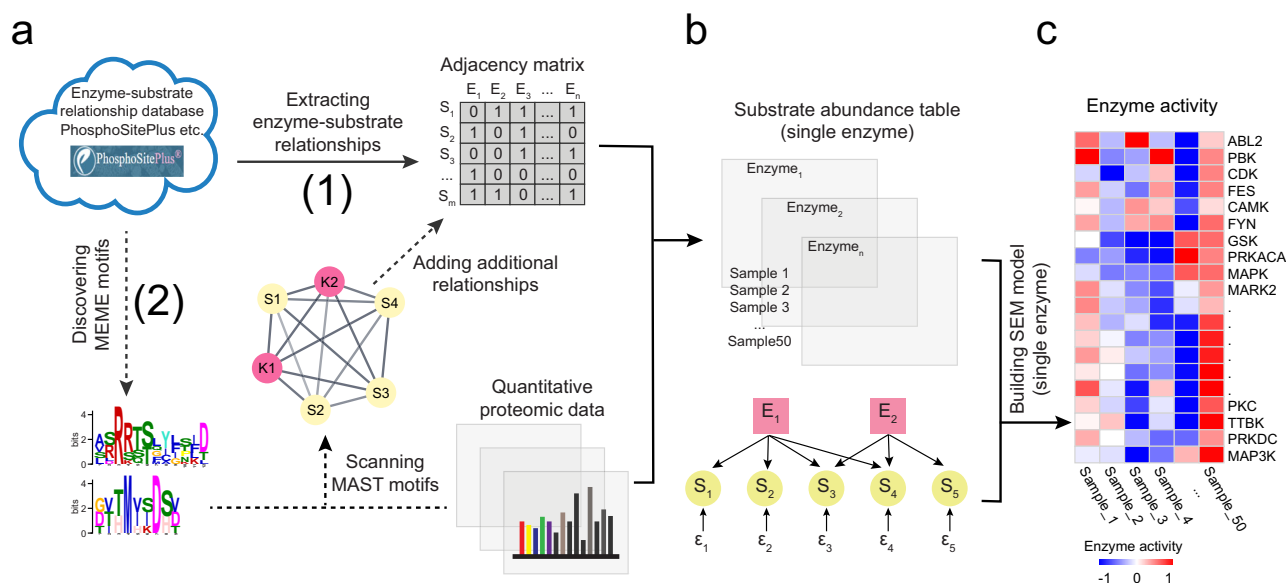


Fig. 1 | Schematic diagram of the JUMPsem pipeline. **a** Construction of enzyme-substrate relationships and discovery of novel enzyme-substrate relationships through motif search. Two approaches are employed: (1) generation of enzyme-substrate relationships from an existing database, indicated by solid lines in the diagram, and (2) motif

searches to predict new enzyme-substrate interactions, denoted by dashed lines. The adjacency matrix and quantitative proteomic data are used to infer kinase activity. **b** Inferring enzyme activity using a structural equation modeling (SEM) algorithm. **c** The output of enzyme activities generated by the JUMPsem pipeline.

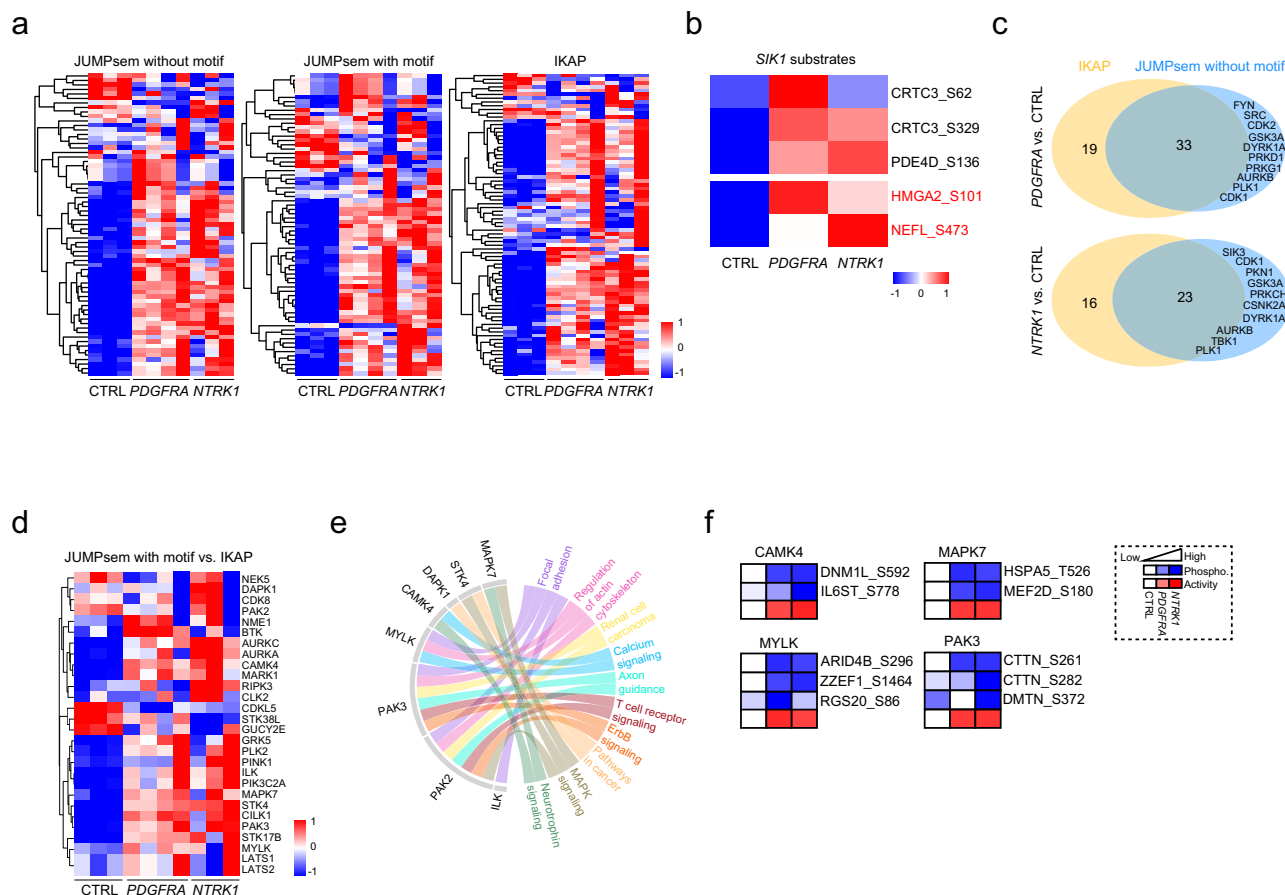


Fig. 2 | Kinase motif-assisted enhancement of kinase-substrate relationship identification. **a** Heatmaps display the kinase activity profiles derived from JUMPsem with and without motif-assisted and IKAP methods within two mouse high-grade glioma (HGG) xenograft models, PDGFRA and NTRK1, contrasted with normal cortex controls. **b** Heatmap illustrating the example *SIK1* substrate abundance from JUMPsem with and without motif assisted. The substrate names marked in red were additionally called out with the assistance of the motif function. **c** Venn

diagram illustrates the overlapping differential kinase identifications for PDGFRA vs CTRL and NTRK1 vs CTRL between JUMPsem with motif assistance and IKAP analyses. **d** Heatmap delineates kinases exclusively detected by JUMPsem with motif assistance, highlighting the method's sensitivity. **e** Circos plot visualizes the pathway enrichment attributable to the motif-assisted JUMPsem, linking kinases to biological processes. **f** Four exemplar protein kinases show differential activity among NTRK1 and PDGFRA-driven glioma models, compared to cortex samples.

The outputs generated by the JUMPsem include the enzyme activity and the affinity coefficient (or connection strength) of each PTM site, representing the estimated effective connectivity with the corresponding enzyme.

Applications of JUMPsem to estimate kinase activity using mouse phosphoproteomic data

We first applied the JUMPsem program to estimate protein kinase activity using the phosphoproteomic data derived from two mouse high-grade glioma (HGG) xenograft models driven by mutated receptor tyrosine kinase (RTK) oncogenes, platelet-derived growth factor receptor alpha (PDGFRA) and fusion genes of the neurotrophic tyrosine receptor kinase 1 (NTRK1), as well as from normal controls (CTRL)²³. Using 45,574 unique phosphopeptides quantified at the peptide FDR of <1%, the JUMPsem program detected a total of 324 substrates with phosphosites that can be found in the PSP database. The JUMPsem program estimated the activity of 67 protein kinases using 177 substrates. By comparing kinase-substrate relationships without motif-assisted prediction, the motif prediction strategy, on average, increases 14.7% (572/3896) kinase-substrate relationships (Fig. 2a). For example, *SIK1* kinase has two substrates (CRTC3 and PDE4D) with three sites being found in the phosphositePlus database. With the motif-assisted function, we found two additional substrates (HMGA2 and NEFL). As expected, these two additional substrates showed a consistent trend, indicating that motif-assisted function selects biologically meaningful substrates instead of random assignment (Fig. 2b).

Our analysis revealed significant alterations in the activity of 33 protein kinases between PDGFRA mice and CTRL are shared both from JUMPsem without motif assisted and IKAP, and 23 kinases between NTRK1 mice and CTRL are shared from these two tools (Fig. 2c). JUMPsem with motif assisted can alternatively infer 28 kinases compared with IKAP (Fig. 2d). These altered kinases were significantly enriched (FDR < 0.01) in cancer-related pathways (Supplementary Data S1a–d), including pathways associated with the checkpoint, and pathways in cancer and glioma. Compared to our previous analysis conducted using the IKAP program, we found a high degree of consistency in the majority of identified pathways between the IKAP and JUMPsem. Motif-assisted strategy allows for the uncovering of previously unknown kinase-substrate relationships, thereby enabling the estimation of the activity of additional kinases. By applying this strategy to the HGG phosphoproteomic data, we further revealed several kinases, such as MAPK7, STK4, DAPK1, CAMK4, MYLK, PAK2, PAK3, and ILK, which were not identified in the previous IKAP analysis. Many of these kinases were implicated in cancer-related pathways, including MAPK and ErbB pathways (Fig. 2e), exhibiting activity alterations without corresponding changes at the protein level (Fig. 2f). This motif-assisted prediction strategy enhances the scope of kinase activity analysis.

Comparison of JUMPsem with other programs applied to human phosphoproteomic data

To assess the performance of JUMPsem, we compared the performance of JUMPsem with IKAP and KSEA, two widely used programs developed for

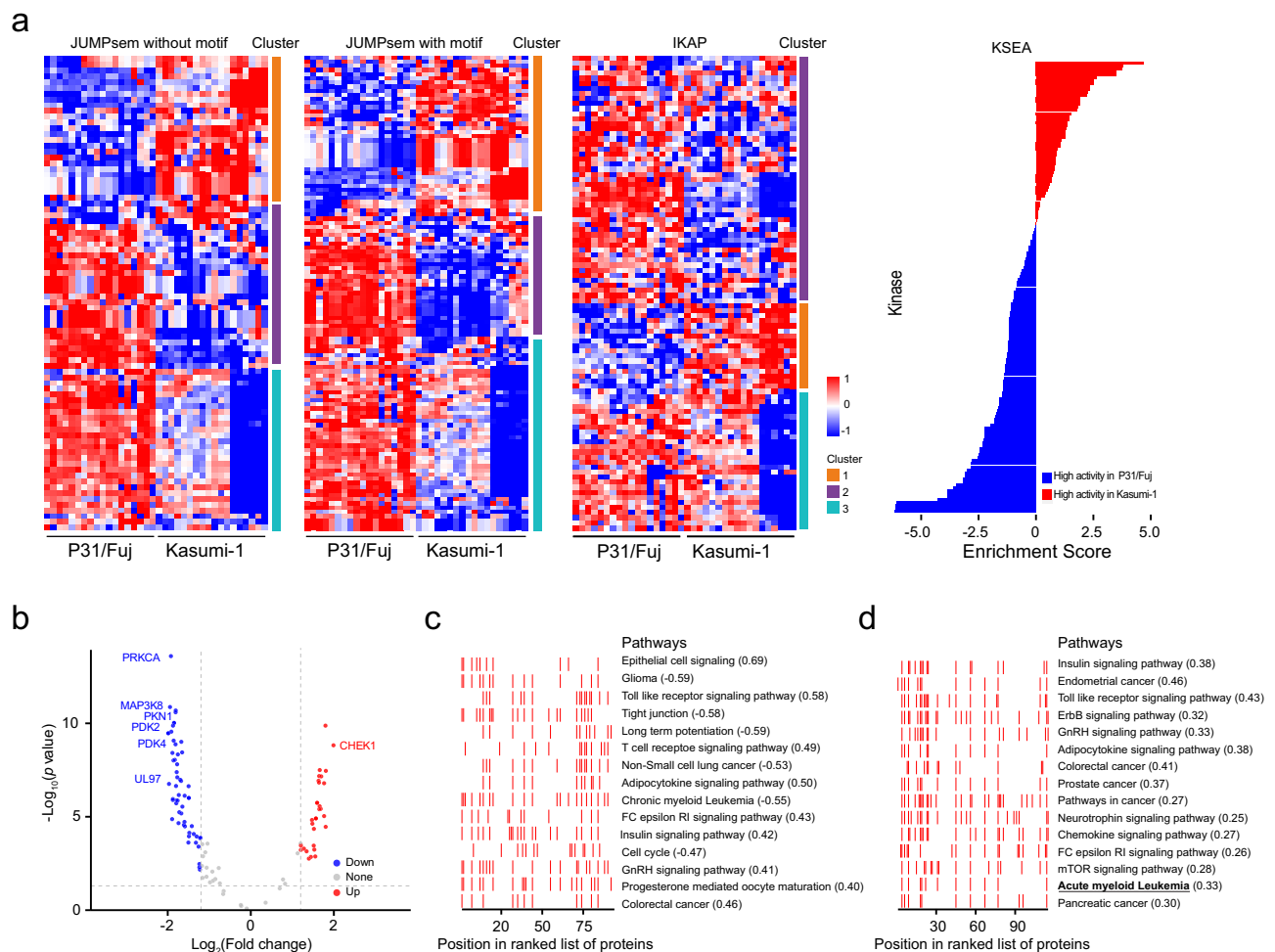


Fig. 3 | Comparison of kinase activity inference methods. a Heatmap depicting the z-score normalized kinase activities between two conditions, P31/Fuj, and Kasumi-1 cell lines, as determined by JUMPsem without motif-assisted search, JUMPsem with motif-assisted search, IKAP, and KSEA. Colors correspond to activity levels: red indicates high kinase activity, and blue indicates low kinase activity. Three clusters (Cluster 1, Cluster 2, and Cluster 3) are annotated to the right of each heatmap, indicating groupings of kinases based on activity patterns. The waterfall plot on the right is the result from KSEA, which we applied to compare with IKAP and our JUMPsem. **b** Volcano plot illustrating the variation in kinase activity between P31/Fuj and Kasumi-1 cell lines, as inferred by JUMPsem with motif search. The x-axis denotes the log₂ fold change, where positive values suggest increased activity in P31/Fuj

relative to Kasumi-1, and negative values denote decreased activity. The y-axis corresponds to the negative logarithm (base 10) of the p-value, with higher values signifying more substantial statistical significance. **c** Bar chart representing pathways enriched in protein kinases with heightened activity in P31/Fuj compared to Kasumi-1 inferred from IKAP. The x-axis indicates the positions of the pathway members in the ranked list of proteins, while the y-axis lists the pathways. The values in parenthesis indicate the enrichment score, quantifying the overrepresentation for each pathway. The order of the protein position is ranked by the fold change score of kinase activity. **d** Bar chart detailing pathways enriched in protein kinases with heightened activity in P31/Fuj relative to Kasumi-1 inferred from JUMPsem with motif assisted, using the same axes format as in (c).

kinase activity estimation, using a human phosphoproteomic dataset. The phosphoproteomic data were generated from the P31/Fuj and Kasumi-1 acute myeloid leukemia (AML) cell lines²⁴. P31/Fuj cells are deficient in PTEN (phosphatase and tensin homolog), whereas Kasumi-1 cells have wild-type PTEN, which is resistant to multiple drugs²⁵. The dataset was originally used to estimate the activity of protein kinases using the KSEA program. Utilizing a total of 4129 phosphorylated peptides quantified across the cell lines, we estimated the activity of protein kinases using JUMPsem, IKAP, and KSEA (Fig. 3a). When comparing the results from IKAP, JUMPsem revealed a cluster that shows a similar kinase activity pattern observed between P31/Fuj and Kasumi-1 cell lines. In addition, JUMPsem identified two unique clusters of kinase activity (Clusters 1 and 2). With a significance threshold set at a p-value < 0.05 and an absolute fold change greater than 1, we identified 75 up-regulated and 31 down-regulated kinases in activity between the P31/Fuj and Kasumi-1 cells (Fig. 3b; Supplementary Data S2). For instance, PRKCA exhibited a notable 7.9-fold decrease in activity in Kasumi-1 compared to P31/Fuj cells. These kinases with differential activity were enriched in the pathways, such as glioma (e.g.,

MAP2K2), Toll-like receptor signaling (e.g., MAP2K2 and MAP2K4), acute myeloid (e.g., MAP2K2) (Fig. 3c, d and Supplementary Data S3a, b).

Performance evaluation of JUMPsem using phosphoproteomic benchmark datasets

To evaluate the performance of JUMPsem, we conducted a comprehensive analysis using benchmark data from two previously published studies^{26,27}, which included 16 phosphoproteomic datasets covering a total of 31 conditions (Supplementary Data S4). We assessed the performance by comparing the precision of JUMPsem and IKAP across various threshold cutoffs. Briefly, for each dataset and condition, we first processed the phosphoproteome, generated a kinase-substrate relationship table, and identified the target kinase (Supplementary Fig. S5a). The average kinase activity was then calculated across all replicates, and the rank of the target kinase was recorded among all inferred kinases for each dataset (Supplementary Fig. S5b). A true positive was defined as a target kinase whose rank was below the threshold cutoff. Precision was finally calculated for performance comparison, which was defined as the ratio of true positives to the

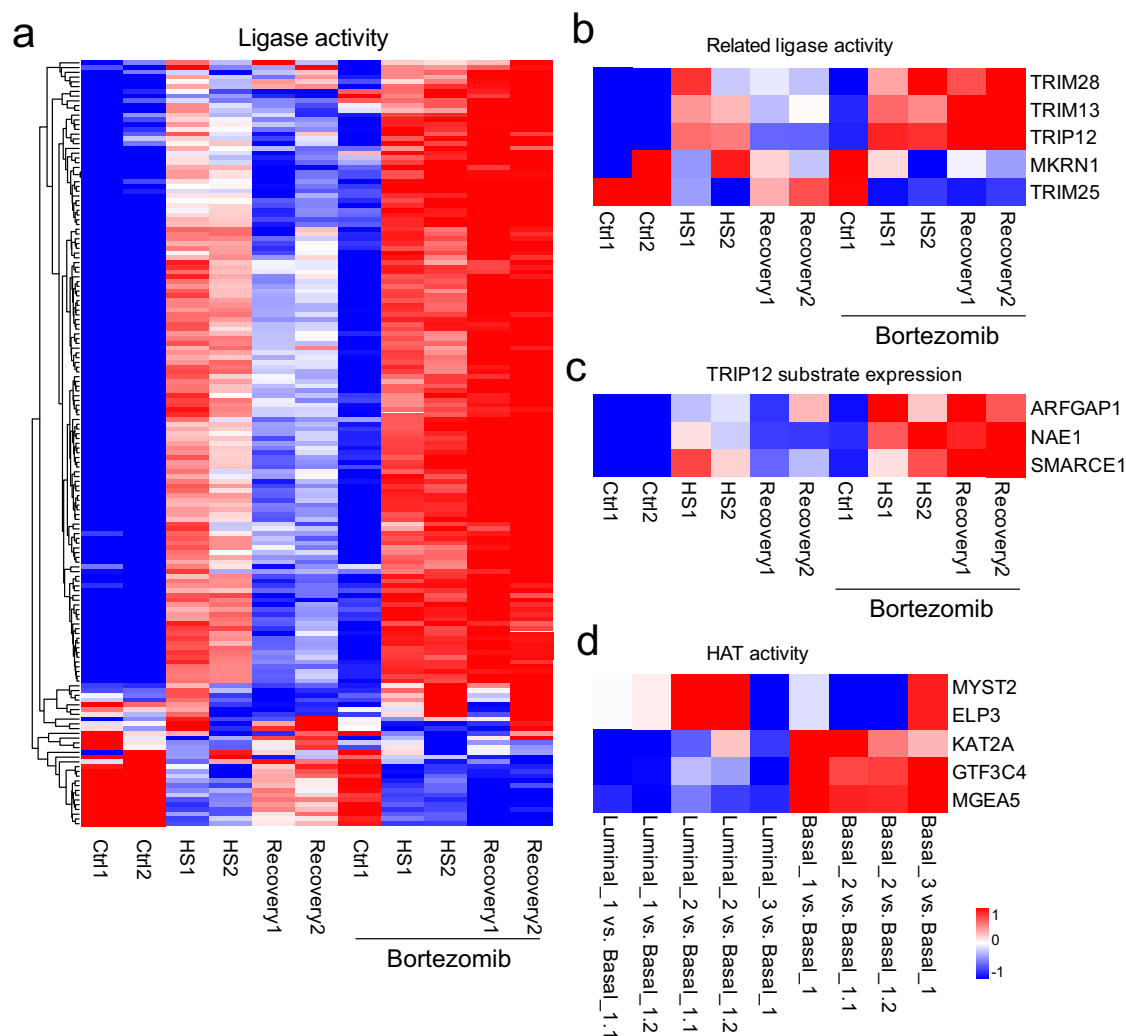


Fig. 4 | Activity profiles of E3 ubiquitin ligases and histone acetyltransferases from ubiquitinome and acetylome data. a Heatmap illustrating the activity profiles of E3 ubiquitin ligases across cells subjected to no stress, heat stress, or arsenite stress, with and without proteasome inhibition (0.5 μ M bortezomib), as indicated by

previously generated ubiquitinomic data. **b** Heatmap of five E3 ligases exhibiting altered activities across five stress conditions. **c** Heatmap depicting consistent expression level changes in three substrates of the TRIP12 ligase. **d** Heatmap representing the activities of histone acetyltransferases in various conditions.

total number of true and false positives. Overall, our analysis demonstrated that JUMPsem slightly outperformed IKAP in terms of precision (Supplementary Fig. S5c).

Inference of activities of E3 ubiquitin ligases and histone acetyltransferases

We next sought to infer the activities of E3 ubiquitin ligases from ubiquitinomic data using the JUMPsem program. The ubiquitinomic data were previously generated from cells under various conditions: unstressed, heat-stressed, or arsenite-stressed, in the presence or absence of 0.5 μ M bortezomib, a proteasome inhibitor²⁸. The ubiquitinomic data were generated by the tandem mass tag (TMT) strategy from samples in response to a 60-min heat stress and following 2 h of recovery, quantifying 16,525 unique ubiquitin-modified peptides, corresponding to 4892 ubiquitinated proteins. The Ubi-Browser database²⁹, which contains a total of 1599 ubiquitinated substrate proteins, corresponding to 408 human E3 enzymes, was used to construct the relationship table of E3 ubiquitin ligases and substrates. The JUMPsem program found 161 E3 ubiquitin ligases and 4892 unique substrate proteins in the database (Fig. 4a). Our findings revealed that 124 unique ubiquitin ligases exhibited significant alterations in their activity between the heat-shock conditions (HS) and the control condition (Ctrl), as determined by a *t*-test (*p*-

value < 0.05), with 80 showing differences between HS and Ctrl, 105 between HS and Ctrl with bortezomib treated (Supplementary Data S5). For example, three ligases (i.e., TRIM28, TRIM13, and TRIP12) were among the ligases upregulated in both HS and Recovery stress conditions compared to the control with bortezomib treated (Fig. 4b). In contrast, two ligases, MKRN1 and TRIM25, displayed a decrease in activity in heat shock with proteasome inhibitor bortezomib (i.e., HS_Bortezomib and Recovery_Bortezomib) compared the control, respectively (Fig. 4b). The ligase activity showed consistency with substrate abundance. For example, TRIP12, especially after bortezomib treatment, a consistent increase in both substrate abundance and ligase activity in the HS and Recovery_Bortezomib groups compared to the control condition (Fig. 4c). To estimate the activity of histone acetyltransferases (HATs), TMT10 quantification techniques for global acetylation profiling, luminal and basal subtype breast cancer xenograft tissue samples were homogenized, lysed, and digested³⁰. A total of 4180 Kac peptides correspond to 3653 Kac sites, and 1804 Kac proteins were quantified. These acetylated proteins were mapped to 6 HATs in the compendium of protein lysine acetylation (CPLA) database, curated for lysine-acetylated substrates with their sites¹⁴ (Fig. 4d). The JUMPsem detected three HATs, including KAT2A, GTF3C4, and MGEA5, showed activities as inhibitors between luminal and basal xenograft samples.

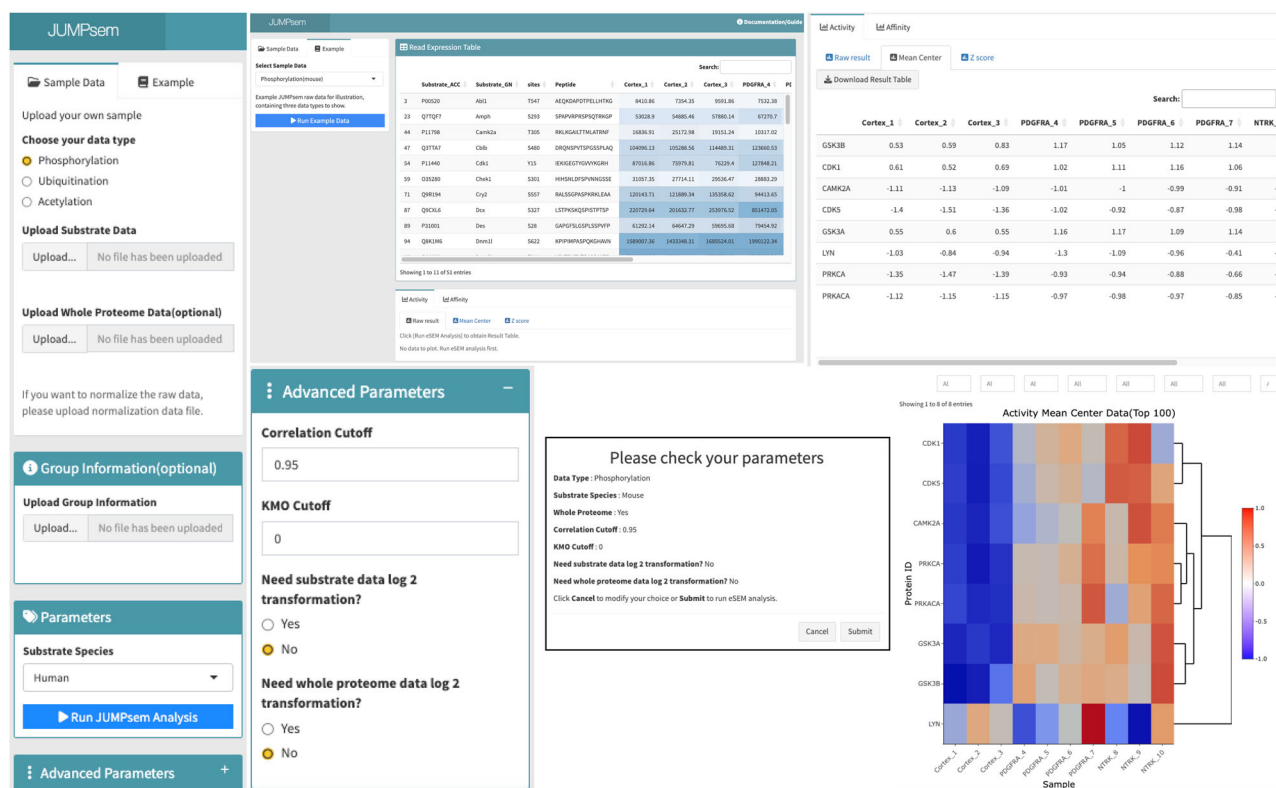


Fig. 5 | JUMPsem R/Shiny web-based platform. The JUMPsem platform workflow encompasses three primary steps: importing data and configuring group parameters, performing JUMPsem analysis, and visualizing enzyme activities.

Modular and scalable R package with shiny web application

JUMPsem is a modular and scalable package, consisting of distinct components designed to process specific types of PTM profiling data. This modular design allows for the future expansion of JUMPsem to incorporate other PTM data, such as methylation profiles. To streamline analysis and visualize results, we have developed a companion R/Shiny application (Fig. 5) that integrates seamlessly with the JUMPsem R package. This web-based JUMPsem R/shiny application can be broadly categorized into three primary steps: loading the raw files, optionally configuring group parameters, and exploring and visualizing kinase analysis. For more comprehensive information and step-by-step guidance, the package vignettes and documentation are available online at the application website (<https://jumpsem.shinyapps.io/JUMPsem/>).

Discussion

Recent advancements in mass spectrometry-based proteomics technologies have enabled us to profile large-scale PTM data. The resulting PTM profiling data are instrumental in unraveling complex signal transduction processes and enzyme activities. Despite the wealth of PTM data generated, the interpretation remains challenging due to a dearth of tools specifically designed to handle and make sense of this complexity. Addressing this critical gap, our study introduces JUMPsem, an innovative computational tool that is capable of analyzing large-scale PTM datasets by facilitating the inference of enzyme activity. The JUMPsem program outperforms existing widely used tools, such as IKAP and KSEA, in terms of the number of enzymes it can analyze and its computational speed. JUMPsem is an effective tool for enzyme activity inference, available as an open-source R package and as an R shiny webtool.

One significant advantage of the JUMPsem program over existing tools for predicting enzyme activity is its integrated motif search strategy. This strategy substantially broadens the potential kinase-substrate interactions that can be identified, going beyond the recorded entries in the PSP database. Despite the fact that proteomics technology is able to detect thousands of modified peptides in a single experiment, a mere fraction of the modified

peptides is cataloged in the public database. For example, less than 10% of the quantified phosphopeptides in an MS-based experiment are represented in the PSP database. Consequently, a vast portion of the data remains underutilized for activity estimation. The JUMPsem program tackles this limitation by augmenting its motif search strategy, thereby extending the detectable range of kinase-substrate relationships beyond the confines of existing databases. This enhanced strategy permits JUMPsem to leverage a more significant segment of the phosphoproteomics data, thus facilitating a more comprehensive analysis of cellular signaling pathways.

Another advantage of the JUMPsem program lies in its utilization of the SEM algorithm, which handles potential measurement errors present in all observed variables. This is particularly important in MS-based phosphoproteomics, where the measurement of modified peptides from large-scale PTM profiling data contains inherent uncertainties and is susceptible to various errors, including quantification measuring errors. In addition, a modified peptide could be incorrectly assigned to a spectrum and a modified site can be misassigned to an adjacent site due to missing informative ions in the spectrum. By accounting for these uncertainties, JUMPsem provides a more robust and reliable estimation of enzyme activities, reflecting the complex reality of biological data.

One limitation of the current JUMPsem program is its design for inferring the activity of individual kinases without considering the interaction of different enzymes within complex signaling networks. These interactions could be incorporated into the SEM model. Most existing computational tools, including the current version of JUMPsem, are tailored to single kinases, which can introduce biases in activity estimation. This bias arises because changes in substrate abundance may be due to the influence of co-active enzyme partners. For instance, fluctuations in the substrate levels of a kinase might be the result of actions by other kinases. Furthermore, the PTM landscape can be influenced by a variety of factors, including kinases relocating within the cell, interactions with other proteins, or competition with de-activation enzymes (e.g., the dynamic between phosphatases and kinases). Such complexities could lead to incorrect estimation of the enzyme activity.

In summary, we introduce the JUMPsem program, an innovative tool designed for inferring protein enzyme activity from PTM profiling data. The modular and scalable JUMPsem package, coupled with its user-friendly web application, makes it a valuable tool for inferring enzyme activity. We have successfully demonstrated the application of JUMPsem to three types of proteomic data—phosphoproteomic, ubiquitomic, and acetylomic—identifying and quantifying enzyme activities, as well as substrate affinities. The utility of JUMPsem in estimating enzyme activity and elucidating cellular signaling pathways marks it as an indispensable tool in the field.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

We processed various public proteomics datasets for our tool-developing analysis, with the example datasets compiled in the software. These include: the mouse phosphoproteomics dataset from GSE114331; the human phosphoproteomics dataset from Casado et al., available at <https://www.mcponline.org/cms/10.1074/mcp.M110.003079/attachment/0541070f-4c56-4031-b0d7-4df51630e974/mmc1.zip>; ubiquitinomic data available from the https://www.science.org/doi/suppl/10.1126/science.abc3593/suppl_file/abc3593_tables8.xlsx; acetylome data available from <https://www.mcponline.org/cms/10.1074/mcp.O114.047555/attachment/3b37fd01-8309-488a-a143-162ab83b68f2/mmc1.zip>. Additionally, we utilized proteomics enzyme-substrate relationship libraries downloaded from PhosphoSitePlus at <https://www.phosphosite.org/staticDownloads>, UbiBrowser 2.0 at http://ubibrowser.bio-it.cn/ubibrowser_v3/home/download and CPLA 1.0 at <http://cpla.biocuckoo.org>. All the processed datasets generated in our study are made available and can be accessed at <https://github.com/Wanglab-UTHSC/JUMPsem/tree/main/data>. Source data are provided with this paper.

Code availability

The source code of JUMPsem can be accessed at <https://github.com/Wanglab-UTHSC/JUMPsem>. The RShiny version code can be accessed at https://github.com/Wanglab-UTHSC/JUMPsem_web. All main figures with related tables and codes are also publicly available in the Zenodo repository at <https://doi.org/10.5281/zenodo.14610849>³¹.

Received: 8 February 2024; Accepted: 14 January 2025;

Published online: 21 January 2025

References

- Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- Kahn, S. E., Hull, R. L. & Utzschneider, K. M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444**, 840–846 (2006).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Bai, B. et al. Proteomic landscape of Alzheimer's disease: novel insights into pathogenesis and biomarker discovery. *Mol. Neurodegener.* **16**, 55 (2021).
- Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28–39 (2009).
- Cohen, P. Protein kinases—The major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309–315 (2002).
- Roskoski, R. Properties of FDA-approved small molecule protein kinase inhibitors: a 2023 update. *Pharm. Res.* **187**, 106552 (2023).
- Whitmarsh, A. J. & Davis, R. J. Analyzing JNK and p38 mitogen-activated protein kinase activity. *Methods Enzymol.* [https://doi.org/10.1016/S0076-6879\(01\)32212-7](https://doi.org/10.1016/S0076-6879(01)32212-7) (2001).
- Shults, M. D., Janes, K. A., Lauffenburger, D. A. & Imperiali, B. A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates. *Nat. Methods* **2**, 277–284 (2005).
- Stewart, E. et al. Identification of therapeutic targets in rhabdomyosarcoma through integrated genomic, epigenomic, and proteomic analyses. *Cancer Cell* **34**, 411–426.e19 (2018).
- Tan, H. et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity* **46**, 488–503 (2017).
- Hornbeck, P. V. et al. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.* **47**, D433–D441 (2019).
- Wang, X. et al. UbiBrowser 2.0: a comprehensive resource for proteome-wide known and predicted ubiquitin ligase/deubiquitinase-substrate interactions in eukaryotic species. *Nucleic Acids Res.* **50**, D719–D728 (2022).
- Liu, Z. et al. CPLA 1.0: an integrated database of protein lysine acetylation. *Nucleic Acids Res.* **39**, D1029–D1034 (2011).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
- Wiredja, D. D., Koyutürk, M. & Chance, M. R. The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **33**, 3489–3491 (2017).
- Kuleshov, M. V. et al. KEA3: improved kinase enrichment analysis via data integration. *Nucleic Acids Res.* **49**, W304–W316 (2021).
- Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
- Bansal, M. et al. Elucidating synergistic dependencies in lung adenocarcinoma by proteome-wide signaling-network analysis. *PLoS One* **14**, e0208646 (2019).
- Ding, H. et al. Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat. Commun.* **9**, 1471 (2018).
- Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
- Rosseel, Y. lavaan: an R Package for Structural Equation Modeling. *J. Stat. Softw.* **48**, 1–36 (2012).
- Wang, H. et al. Deep multiomics profiling of brain tumors identifies signaling networks downstream of cancer driver genes. *Nat. Commun.* **10**, 3718 (2019).
- Casado, P. et al. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* **6**, rs6 (2013).
- Casado, P. & Cutillas, P. R. A self-validating quantitative mass spectrometry method for assessing the accuracy of high-content phosphoproteomic experiments. *Mol. Cell. Proteom.* **10**, M110.003079 (2011).
- Hernandez-Armenta, C., Ochoa, D., Gonçalves, E., Saez-Rodriguez, J. & Beltrao, P. Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics* **33**, 1845–1851 (2017).
- Crowl, S., Jordan, B. T., Ahmed, H., Ma, C. X. & Naegle, K. M. KSTAR: an algorithm to predict patient-specific kinase activities from phosphoproteomic data. *Nat. Commun.* **13**, 4283 (2022).
- Maxwell, B. A. et al. Ubiquitination is essential for recovery of cellular activities after heat shock. *Science* **372**, eabc3593 (2021).
- Li, Y. et al. An integrated bioinformatics platform for investigating the human E3 ubiquitin ligase-substrate interaction network. *Nat. Commun.* **8**, 347 (2017).
- Svinkina, T. et al. Deep, quantitative coverage of the lysine acetylome using novel anti-acetyl-lysine antibodies and an optimized proteomic workflow. *Mol. Cell. Proteom.* **14**, 2429–2440 (2015).
- Kong, D. et al. A computational tool to infer enzyme activity using post-translational modification profiling data. Zenodo <https://doi.org/10.5281/zenodo.14610849> (2025).

Acknowledgements

This work was supported by the NIH grants R01DA056523, RF1AG072703, R01NS123456, and R01DK130913 (to X.W.) and RF1AG064909 and RF1AG068581 (to J.P.). The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Author contributions

Conceptualization: D.K., J.P. and X.W. Methodology: D.K. and X.W. Software: D.K. Investigation: D.K., A.Z., and X.W. Data curation: D.K. A.Z., L.L., Z.Y., Y.F., L.W., A.M., and X.W. Writing—original draft: D.K. and X.W. Writing—review and editing: D.K., A.H., J.P. and X.W. Funding acquisition: X.W. and J.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07548-4>.

Correspondence and requests for materials should be addressed to Junmin Peng or Xusheng Wang.

Peer review information *Communications Biology* thanks Mariano Alvarez and Dong Li for their contribution to the peer review of this work. Primary Handling Editors: Tuan Anh Nguyen and Dario Ummarino.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025