DORSSAA: Drug-target interactOmics Resource based on Stability/Solubility Alteration Assay

Ehsan Zangene[1], Elham Gholizadeh[1], Amir Ata Saei[2], Mathias Wilhelm[3], Mohieddin Jafari[1,*]

(1) Department of Biochemistry and Developmental Biology, University of Helsinki, Helsinki, Finland

(2) Department of Microbiology, Tumor and Cell Biology at Karolinska Institutet, Stockholm, Sweden

(3) Department of Biological Science of Technical University of Munich, Munich, Germany

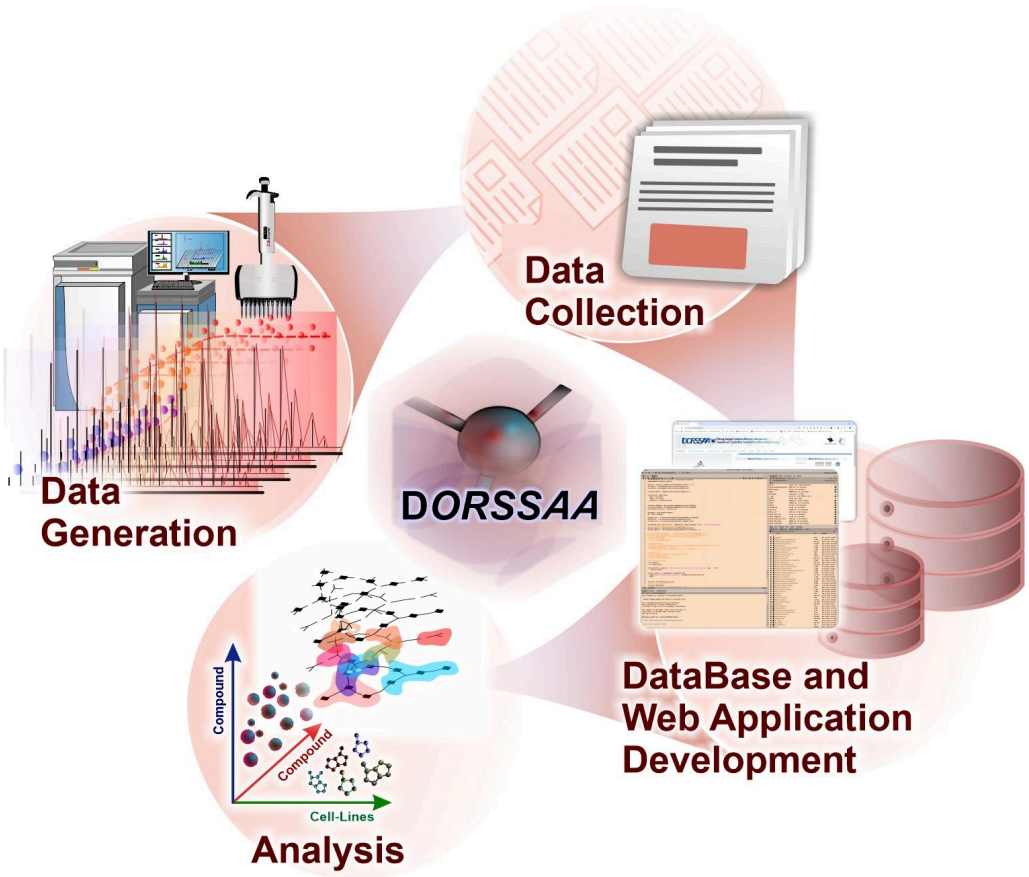
* Corresponding author:

Mohieddin Jafari,

mohieddin.jafari@helsinki.fi

ABSTRACT

Advancements in high-throughput techniques such as Thermal Proteome Profiling (TPP) and the high-throughput Proteome Integral Solubility Alteration (PISA) assay have revolutionized our understanding of drug-protein interactions. Despite these innovations, the absence of an integrative platform for cross-study analysis of stability and solubility alteration data represents a significant bottleneck. To address this gap, we introduce DORSSAA (Drug-target interactOmics Resource based on Stability/Solubility Alteration Assay), an interactive and expandable web-based platform for the systematic analysis and visualization of proteome stability and solubility alteration assay datasets. Currently, DORSSAA features 480,456 records spanning 37 cell lines and organisms, 39 compounds, and 40,004 potential protein targets. Through its user-friendly interface, the resource supports comparative drug-protein interaction analysis and facilitates the discovery of actionable therapeutic targets. Using two case studies of the DHFR-methotrexate interaction and drug-target interaction in the case of combinatorial therapy in Leukemia cell lines, we demonstrate DORSSAA's utility in identifying protein-drug interactions across experimental conditions. This resource empowers researchers to accelerate drug discovery and enhance our understanding of protein behavior.



Introduction

Drug discovery is a multidisciplinary field encompassing medicine, biotechnology, and pharmacology, aimed at identifying novel therapeutic compounds. Despite significant advances, a major challenge remains in the precise characterization of drug targets and off-targets within complex cellular environments. This challenge is especially pronounced in oncology, where identifying cancer vulnerabilities and dependencies, such as synthetic lethality relationships, is vital for developing precision therapies[1,2]. This limitation is compounded by the absence of direct, systematic methods to monitor the effects of drugs on proteins within cells and tissues[3]. Traditional evaluations by regulatory bodies such as the FDA, including absorption, distribution, metabolism, and excretion (ADME) studies, and animal toxicity assessments, often fall short of fully elucidating a drug's mechanism of action (MoA)[4]. From a drug MoA point of view, identifying effective drug targets is challenging in drug development. As a result, many drug candidates fail during late-stage clinical trials, (phase II trials), leading to resource wastage and underscoring the urgent need for innovative strategies to enhance drug development[5].

While small molecule modulators hold promise as therapeutic agents, their diverse effects on protein targets necessitate comprehensive profiling to better understand their molecular MoA. Although transcriptomic approaches like RNA sequencing and projects such as the Connectivity Map (CMAP) have been instrumental in mapping drug action[6]. However, given that most FDA-approved drugs target proteins and the weak correlation between protein and RNA levels, protein-level changes are likely the most relevant readout of compound action[7–10].

Recent innovations in solubility and stability alteration assays have revolutionized the study of drug-protein interactions[11,12]. These techniques including cellular thermal shift assay (CETSA), thermal proteome profiling (TPP), and proteome integral solubility alteration (PISA) assay leverage changes in protein stability or solubility upon ligand binding or environmental perturbations, offering an unbiased, high-throughput approach to probe drug interactions at the proteome level. By subjecting samples to varying conditions, these methods uncover critical insights into protein behavior and drug interactions across diverse biological contexts[13,14]. Due to its numerous advantages, this approach has become widely used in both drug discovery and chemical proteomic research.

Here, we present **DORSSAA** (Drug-target interactOmics Resource based on Stability Solubility Alteration Assay; https://dorssaa.it.helsinki.fi/), a comprehensive resource for the systematic analysis and visualization of solubility and stability alteration datasets. DORSSAA integrates high-throughput experimental data encompassing 480,456 records from 37 cell lines and organisms, 39 compounds, and 40,004 proteins, providing an unprecedented dataset for drug interactomics. By consolidating this data into an intuitive, user-friendly web application, DORSSAA enables researchers to explore drug-target interactions across diverse experimental conditions. This resource not only facilitates cross-validation of findings but also accelerates the identification of actionable drug targets, enhancing the potential for translational drug discovery.

Materials and Methods

Collection of Proteomics Data

We compiled publicly available proteomics datasets from diverse sources to conduct a comprehensive analysis of drug-target interactions. This study incorporated 24 datasets, including experiments such as TPP, PISA assay, and Improved PISA (i-PISA), among others. These datasets encompassed 39 distinct chemical compounds tested across 37 unique cancer cell lines or species. Detailed information on the datasets and experimental conditions is provided in Supplementary Table S1 and Figure 1. For each identified protein target, pairwise associations between drugs and target stability or destabilization were systematically evaluated by calculating fold changes or differences and corresponding P-values across all datasets.

In addition to the publicly available datasets, this study also integrates our in-house CoPISA dataset[15], which leverages advanced proteomic tools to characterize drug-protein interactions in combinatorial therapies, providing novel insights into cooperative mechanisms that are unattainable with single-agent approaches.

Data Normalization and Handling Missing Values

To ensure consistency across datasets, we standardized protein identifiers and column naming conventions. Separator characters within column names, introduced by varying study-specific formats, were uniformly replaced with underscores for clarity. Protein names were associated with UniProt IDs and HUGO gene symbols, with UniProt IDs prioritized. When unavailable, HUGO gene symbols were used, and corresponding UniProt IDs were identified using the *biomaRt* R package[16]. A distinctive cloud icon (☘) was incorporated alongside UniProt IDs to enhance user recognition. Interaction records involving multiple gene symbols were systematically split into distinct entries within the database.

To address variability in -log10(p-value) measurements for drug-target interactions, stemming from differing study methodologies, normalization procedures were applied. Variability, visualized through volcano plots in the DORSSAA web application, highlights the dependency of p-values on study-specific conditions. This feature allows users to evaluate the reliability and contextual significance of reported interactions. For datasets containing multiple experimental replications, an additional column was included in the reported data table, indicating the replication status of each interaction. This provides an extra layer of insight into the reproducibility of the findings.

Protein Query Database Interface

The Protein Query tab provides a versatile platform for users to investigate compounds treated within specific cellular contexts or organisms, enabling the identification of individual drug-target interactions or comprehensive database exploration. Query outputs are presented in multiple formats, including bar charts, volcano plots, and detailed data tables containing interaction-specific

information. Users can further refine these results based on the significance thresholds of reported p-values.

Additionally, the interface facilitates cross-referencing of protein identities with UniProt records and provides direct access to external databases, such as DTC[17], Micha, and Drug Target Interact[18], for extended information on drug-target interactions[17,19].

Cell Lines Query Database Interface

The Cell Line Query tab offers a powerful platform for investigating critical biological questions by enabling users to explore the scope and extent of drug-protein interactions across various cellular contexts. By integrating insights from the Protein Query tab, users can evaluate which cancer cell lines or organisms are most conducive to drug-protein interactions and which are less responsive. This functionality provides valuable insights into pan-cancer and cancer-specific interactions, supporting the development of personalized therapeutic strategies. To enhance user experience and analysis capabilities, the interface includes advanced tools and features, such as interaction filtering based on p-values, UniProt protein annotations, external checks with UniProt, and cross-references to external drug-target interaction databases. These resources ensure comprehensive exploration and interpretation of interaction data within relevant biological contexts.

Additionally, the interface incorporates a melting points distribution tool sourced from the Meltome database[20]. This tool allows users to visualize the melting point diversity of queried proteins across different contexts. Using box plots separated by cell line and a combined box plot for all cell lines, researchers can gain critical insights into melting point variations, providing a valuable perspective when designing experiments.

Compound query database interface

The Compound Query tab addresses critical aspects of drug repurposing by enabling users to investigate how a single protein can be targeted by multiple drugs. This feature highlights drugs with potential therapeutic applications against a specific protein within a defined cellular or organismal context, thus fostering innovative approaches to drug repurposing and therapy design.

Similar to the Cell Line Query tab, this interface includes a robust suite of analytical tools, including filtering interactions by p-values, accessing UniProt protein descriptions, performing external checks with UniProt, and verifying drug-target interactions through external databases. Additionally, we provided the same functionality mentioned in the Cell Line Query tab for protein melting points, allowing users to access insights into melting point diversity across different cell lines and contexts.

Protein Set Enrichment Tab

To enhance functional insights, the platform includes built-in tools for gene set enrichment analysis (GSEA) and singular enrichment analysis (SEA) based on Gene Ontology (GO) annotation[21].

These tools enable the investigation of potential shared functional characteristics among proteins associated with specific compounds. The analysis is rooted in the hypothesis that proteins targeted by a given chemical may share molecular functions, biological processes, or cellular compartments as defined by GO terms. Results of enrichment analyses are visualized through treemaps and network diagrams, offering intuitive representations of interrelated GO terms with similar semantic features.

This framework further supports the hypothesis that compounds targeting distinct protein groups may exhibit functional similarities, reflected in shared gene ontologies. By incorporating semantic similarity tools, users can quantitatively assess and visualize these relationships, providing valuable insights into the functional and mechanistic overlap among drug targets.

Database Implementation

The DORSSAA backend server was developed using the R programming language (v.4.3.2) in conjunction with the RStudio Shiny framework to provide an interactive web-based user interface. SQL database functionality was integrated via the DBI R package, enabling efficient data management and querying. JavaScript code was incorporated on both the front-end and server sides to enhance functionality and interactivity. Statistical analyses and data processing were conducted entirely in R. DORSSAA was rigorously tested for compatibility and performance across multiple widely-used web browsers, including Google Chrome (preferred), Firefox, and Apple Safari. The system ensures seamless user interaction regardless of the chosen platform. The design and development workflow for DORSSAA is illustrated in Figure 1.

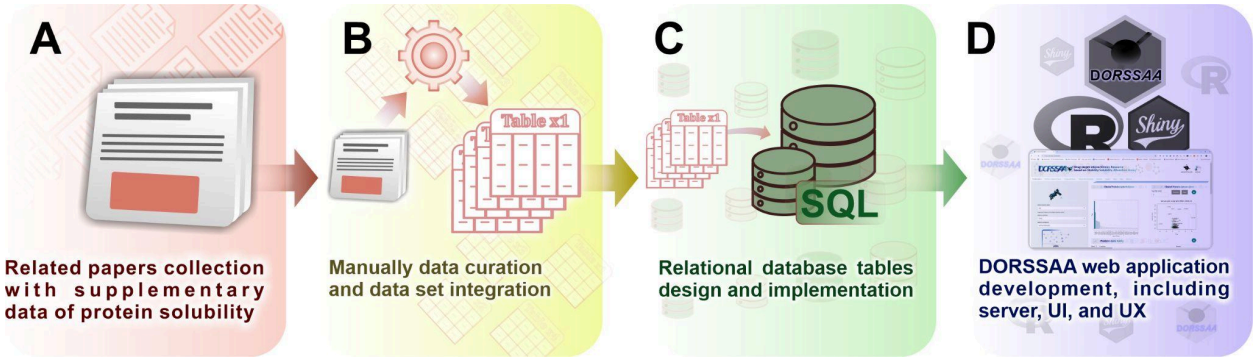


Figure 1: Schematic representation of the DORSSAA workflow, including:, A) Data collection from relevant publications., B) Manual data curation and integration into datasets., C) Relational database schema design. D) Implementation and development of the DORSSAA web application.

Database Content

DORSSAA contains a comprehensive and diverse dataset tailored to the needs of researchers studying drug-protein interactions:

A. **Number of Records**: The core "Drug-Protein Interaction" table includes 480,456 entries, representing interactions derived from 18 distinct cell lines and organisms, 37 unique compounds, and 40,004 proteins (by 30 January 2025).
B. **Data Accessibility**: Researchers can interact with the data through the DORSSAA web application, an intuitive and interactive web-based interface available at https://dorssaa.it.helsinki.fi/. The interface allows users to retrieve results for any drug-target interaction based on proteome solubility assay data.
C. **Ongoing Updates**: The database is updated regularly as new publications with publicly available TPP datasets are released. This ensures that DORSSAA remains a current and reliable resource for the scientific community.
D. **Long-term Availability**: The graphical user interface (GUI) offers free access to researchers and is committed to indefinite availability, making DORSSAA a dependable and enduring tool for drug discovery and development investigations.

Web interface

The DORSSAA web application offers an intuitive and interactive platform for data retrieval, visualization, and export (Figure 2). The interface is structured around three main functional tabs: Protein Query, Cell Line Query, and Compound Query. Users can initiate their queries by selecting options from three dropdown menus: UniProt_ID, Cell Line, and Protein. Upon making these selections, the main panel dynamically generates a bar chart and a volcano plot, visually representing the queried data. Beneath the visual outputs, a detailed data table is displayed, summarizing all relevant interactions with comprehensive annotations and metrics (Figure 2). The visual and tabular outputs enable users to delve into drug-target interactions, assess protein-level changes, and extract meaningful insights into underlying molecular mechanisms. By leveraging these interactive tools, researchers can seamlessly analyze complex datasets and obtain actionable results. The DORSSAA web application thus serves as a robust resource for exploring and interpreting drug-target interaction data, fostering advancements in understanding molecular processes and supporting various research endeavors.

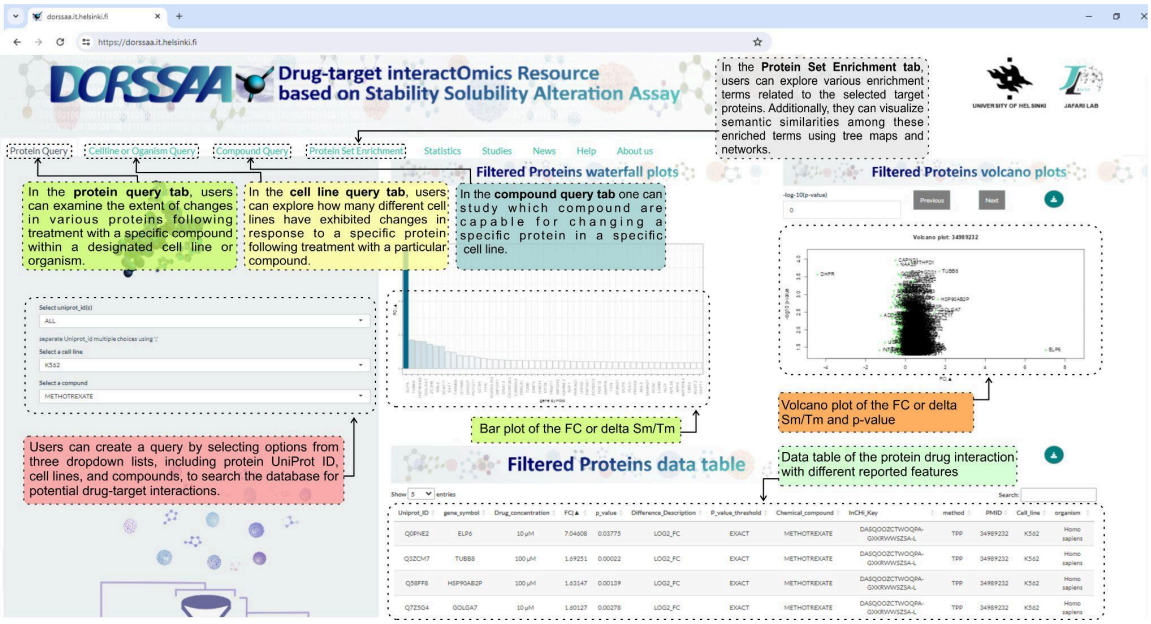


Figure 2: The DORSSAA web application showcases an intuitive interface divided into three primary tabs: Protein Query, Cell Line Query, and Compound Query. Users can initiate queries by selecting options from dropdown menus for UniProt_IDs, cell lines, and compounds. The main panel features visualizations such as bar charts and volcano plots, offering insights into drug-target interactions and protein-level changes. Below the visualizations, a detailed data table provides comprehensive information on all identified interactions, supporting efficient data exploration and in-depth analysis.

Case Studies

Case Study 1: Exploring Methotrexate's Protein Targets and Mechanism of Action in K562 Cells Using DORSSAA

To evaluate the accuracy and relevance of DORSSAA's data, we conducted an in-depth analysis focusing on methotrexate's protein targets within the K562 cell line. Applying stringent filtering criteria, including a p-value threshold of 0.01, we identified statistically significant and biologically relevant interactions.

Using DORSSAA's Protein Query tool, we generated a detailed list of methotrexate-associated proteins. Cross-referencing with UniProt and other external databases revealed a high alignment, demonstrating the reliability of DORSSAA's outputs. Subsequent GO enrichment analysis identified numerous proteins with shared molecular functions, highlighting 17 distinct categories. For example, 101 proteins were enriched in RNA binding, while 119 proteins were involved in nuleic acid binding .

Among the 736 protein targets identified, Dihydrofolate Reductase (DHFR) (UniProt ID: P00374) emerged as a top target. DHFR exhibited the highest log2-fold change (log2-FC), with a highly significant p-value, underscoring its strong association with methotrexate in K562 cells (Figure 3A). DHFR is the cognate target of methotrexate. DHFR is known for its role in the biosynthesis of amino acids and folic acid, reducing dihydrofolate to tetrahydrofolate via the nicotinamide dinucleotide phosphate cofactor. Its therapeutic relevance in cancer and bacterial infections is well-documented. Interestingly, in DORSSAA, DHFR was targeted by five additional drugs (e.g., staurosporine, SNS032, raltitrexed, and panobinostat) across different studies, as shown in the Compound Query tab (Figure 3B).

To explore methotrexate's MoA, biological process enrichment analysis revealed significant terms, including its inhibition of transmethylation reactions. This interaction was further validated across multiple studies using techniques like TPP and Single-tube Thermal Proteome Profiling (STPP), spanning cell lines such as K562 and 293T. Notably, while K562 cells are derived from chronic myelogenous leukemia, 293T cells originate from embryonic kidney cells, suggesting a degree of universality in DHFR-methotrexate interactions. Furthermore, DORSSAA enabled the identification of drug repurposing opportunities for targeting DHFR using other compounds, highlighting its potential as a target for combinatorial and personalized therapeutic strategies (Figure 3C).

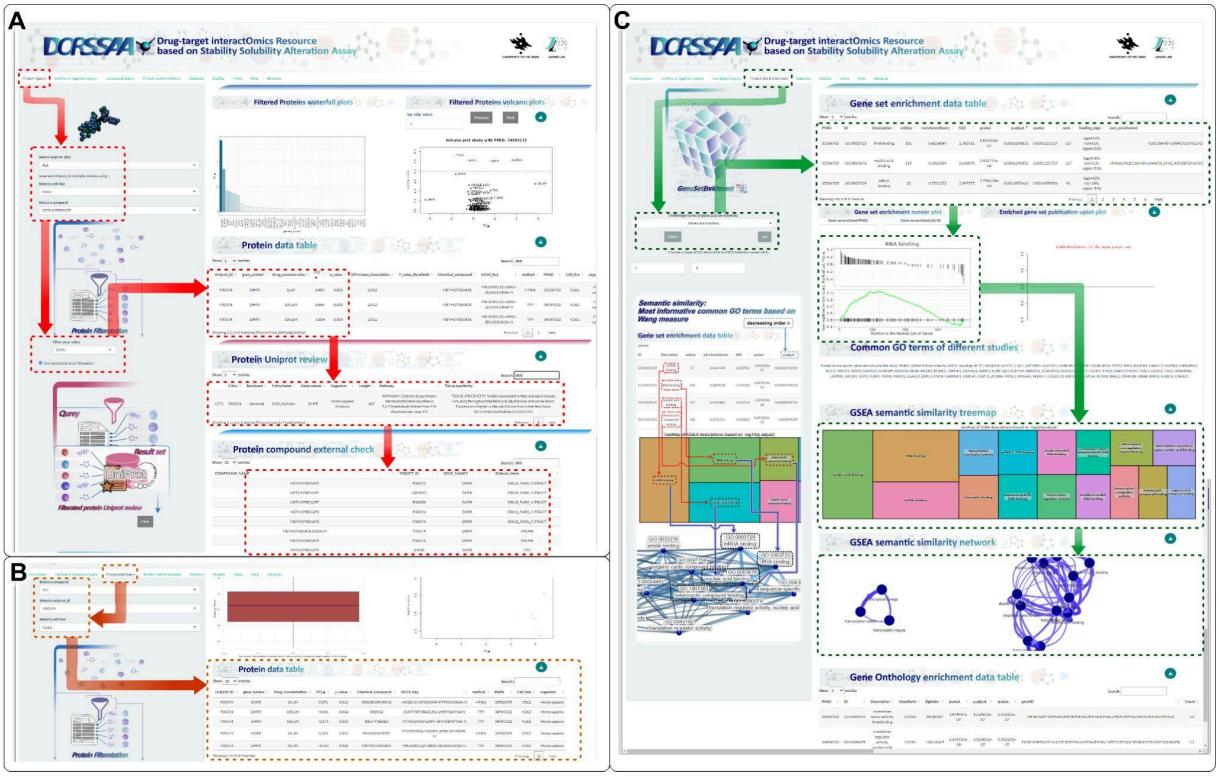


Figure 3. Methotrexate-Associated Proteins and DHFR Target Identification Using DORSSAA (A) Protein Query analysis identified 736 methotrexate-associated proteins, with DHFR showing the highest log2-fold change and significance. (B) DHFR was also targeted by five additional drugs, as shown in the Compound Query tab. (C) Enrichment analysis and cross-referenced studies confirmed DHFR's central role in methotrexate's MoA and highlighted opportunities for drug repurposing and combinatorial therapy.

Case Study 2: Drug-Protein Interactions in Molm16 Cells Investigated Using DORSSAA

A complete submatrix inside the DORSSAA dataset (6 compounds × 4 cell lines) is matched to our recent studies of combinatorial therapy on 4 different AML cell lines with 4 single drugs and 2 combinations [15,22,23]. In this case study, we leveraged DORSSAA's unique analytical capabilities to investigate context-specific drug-target interactions and their combinatorial effects.

To begin, we analyzed the effects of sapanisertib on the Molm16 cell line. In the Protein Query tab, significant targets were identified by applying a p-value threshold of 0.01. Among the resulting targets, RASAL3 (UniProt ID: Q86YV0) stood out with a log-fold change of 0.32 and a p-value of 0.0079, indicating significant drug-induced alterations compared to untreated controls (Figure 4A) .

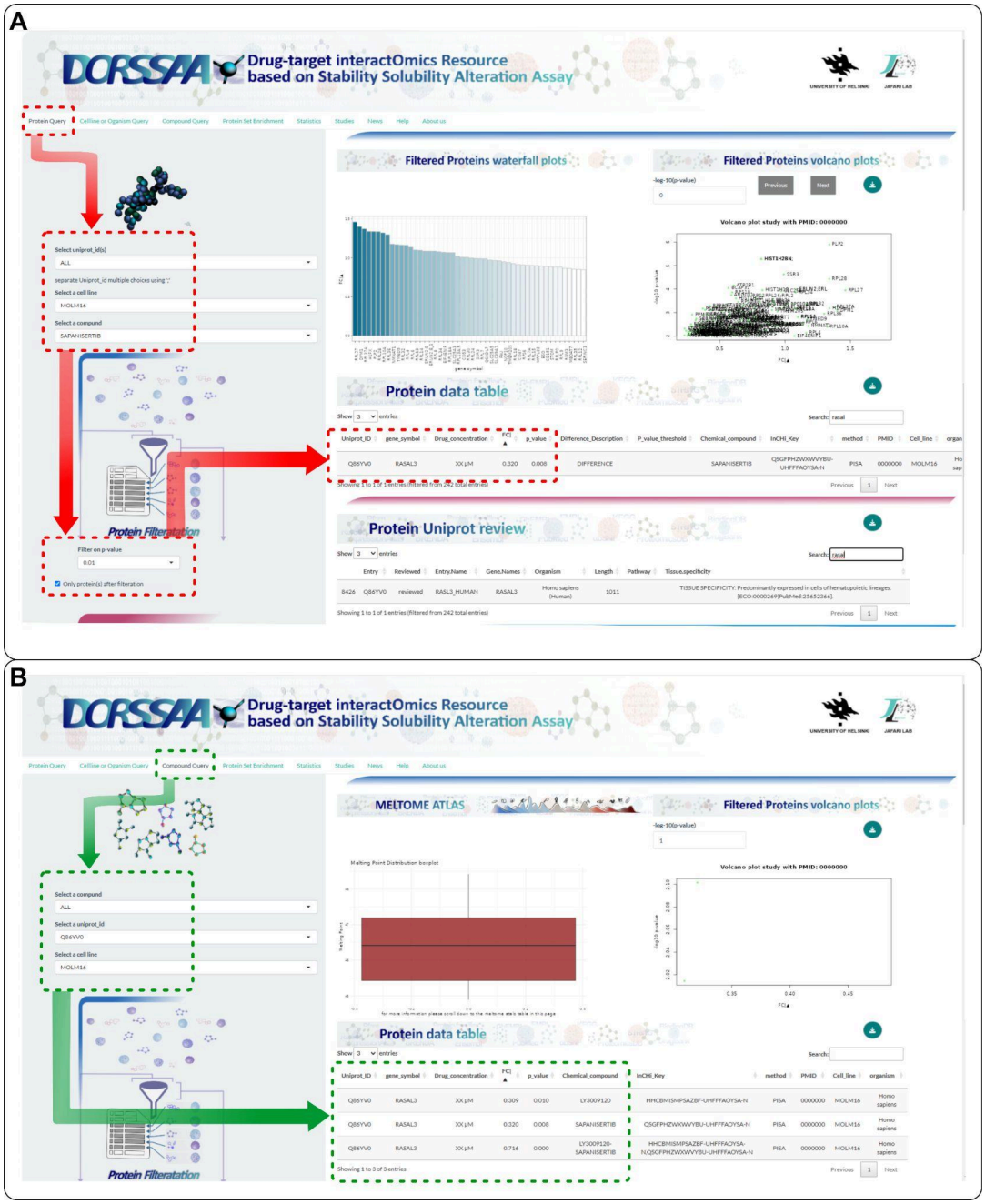


Figure 4: Drug-target Interactions in Molm16 Cells Using DORSSAA
(A) Sapanisertib treatment in Molm16 cells identified RASAL3 (UniProt ID: Q86YV0) as a significant target, with a log-fold change of 0.32 and p-value of 0.0079. (B) LY3009120 also targeted RASAL3 (log-fold change 0.309, p-value 0.0096), and their combination enhanced the effect (log-fold change 0.71, p-value 1.59×10^{-5}).

In the Cell Line/Organism Query tab, we examined the specificity of this interaction across various cell lines and cancer subtypes. The analysis confirmed that the interaction was significant

only in the Molm16 cell line, suggesting a context-specific relationship that could inform precision therapeutic strategies.

Further analysis in the Compound Query tab identified LY3009120 as an additional compound targeting RASAL3, with a log-fold change of 0.309 and a p-value of 0.0096. The combinatorial use of LY3009120 and sapanisertib demonstrated an amplified interaction with RASAL3, yielding a log-fold change of 0.71 and a p-value of 1.59×10^{-5} (Figure 4B).

This phenomenon, reported here for the first time, demonstrates that the combinatorial therapy of LY3009120-sapanisertib exhibits a stronger synergistic effect on the fold change and p-value compared to individual treatments. Such clear synergy highlights the potential for combinatorial strategies in targeted therapy. Importantly, discovering these effects is possible only through the powerful visualization and analysis capabilities of DORSSAA, which integrates diverse datasets and provides biologists with advanced tools to uncover novel relationships and actionable insights.

This comprehensive case study underscores DORSSAA's strength in enabling the identification and exploration of both individual and combinatorial drug-target interactions, providing valuable guidance for targeted drug development and therapeutic innovation.

Conclusion

DORSSAA represents a groundbreaking resource in drug discovery and protein research addressing persistent challenges in assessing cellular-level target engagement and off-target binding. By integrating advanced techniques TPP and PISA, DORSSAA leverages the power of mass spectrometry-based proteomics to systematically characterize drug-protein interactions, offering an innovative solution for the field.

The main objective of DORSSAA is to accelerate drug discovery and enhance our understanding of protein behavior under different conditions. Its comprehensive database compiles differential proteome solubility data sets from various sources, offering a rich and diverse proteomic landscape for researchers to explore. This user-friendly web application enhances this accessibility, offering intuitive tools for visualizing, analyzing, and interpreting drug-target interactions, thus bridging the gap between complex datasets and actionable insights.

One of DORSSAA's core strengths lies in its ability to answer critical biological questions. Researchers can investigate how specific proteins respond to compound treatments across different cell lines or organisms, while also exploring broader patterns of change under varying experimental contexts. Furthermore, the compound query functionality enables users to identify drugs that may affect particular proteins, supporting drug repurposing efforts and the development of combination therapies.

The robust technical foundation of DORSSAA, built on R programming, SQL database functions, and JavaScript, ensures a seamless user experience. Rigorous testing across widely used web browsers guarantees compatibility and reliable performance. The database itself boasts extensive records, featuring data on 37 distinct cell lines and organisms, 39 unique compounds, and 40,004 proteins, making it a highly valuable, versatile and expandable resource.

To remain relevant in the rapidly evolving landscape of proteomics and drug discovery, DORSSAA is committed to continuous updates. By promptly incorporating newly published, publicly available datasets on differential proteome solubility, it ensures that users always have access to the most up-to-date information. This dedication to staying current reinforces DORSSAA's role as an essential tool for researchers striving to address complex challenges in drug discovery and protein science.

Acknowledgment

We would like to acknowledge Salma Ibourki for her assistance with data collection during her summer visit to our lab.

Reference

1. Zangene, E., Marashi, S.-A. & Montazeri, H. SL-scan identifies synthetic lethal interactions in cancer using metabolic networks. *Sci Rep* **13**, 15763 (2023).
2. Barneh, F. *et al.* Integrated use of bioinformatic resources reveals that co-targeting of histone deacetylases, IKBK and SRC inhibits epithelial-mesenchymal transition in cancer. *Brief Bioinform* **20**, 717–731 (2019).
3. Seashore-Ludlow, B., Axelsson, H. & Lundbäck, T. Perspective on CETSA Literature: Toward More Quantitative Data Interpretation. *SLAS Discov* **25**, 118–126 (2020).
4. Brown, D. G. & Wobst, H. J. Opportunities and Challenges in Phenotypic Screening for Neurodegenerative Disease Research. *J. Med. Chem.* **63**, 1823–1840 (2020).
5. Phaneuf, C. G. *et al.* Experimental strategies to improve drug-target identification in mass spectrometry-based thermal stability assays. *Commun Chem* **6**, 64 (2023).
6. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452.e17 (2017).
7. Saei, A. A. *et al.* ProTargetMiner as a proteome signature library of anticancer molecules for functional discovery. *Nat. Commun.* **10**, 5715 (2019).
8. Mitchell, D. C. *et al.* A proteome-wide atlas of drug mechanism of action. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01539-0.
9. Gygi, S. P., Rochon, Y., Robert Franza, B. & Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Molecular and Cellular Biology* vol. 19 1720–1730 Preprint at https://doi.org/10.1128/mcb.19.3.1720 (1999).
10. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387–402.e16 (2020).
11. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).

12. Gholizadeh, E. *et al.* Identification of Celecoxib-Targeted Proteins Using Label-Free Thermal Proteome Profiling on Rat Hippocampus. *Mol. Pharmacol.* **99**, 308–318 (2021).
13. Gaetani, M. *et al.* Proteome Integral Solubility Alteration: A High-Throughput Proteomics Assay for Target Deconvolution. *J. Proteome Res.* **18**, 4027–4037 (2019).
14. Perrin, J. *et al.* Identifying drug targets in tissues and whole blood with thermal-shift profiling. *Nat. Biotechnol.* **38**, 303–308 (2020).
15. Gholizadeh, E. *et al.* Exploring mechanisms of action in combinatorial therapies through solubility alterations: Advancing AML treatment. *bioRxiv* (2024) doi:10.1101/2024.11.08.618644.
16. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* **4**, 1184–1191 (2009).
17. Tanoli, Z. *et al.* Drug Target Commons 2.0: a community platform for systematic analysis of drug-target interaction profiles. *Database (Oxford)* **2018**, 1–13 (2018).
18. Chen, X. *et al.* Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform* **17**, 696–712 (2016).
19. Tanoli, Z. *et al.* Minimal information for chemosensitivity assays (MICHA): a next-generation pipeline to enable the FAIRification of drug screening experiments. *Brief Bioinform* **23**, (2022).
20. Jarzab, A. *et al.* Meltome atlas-thermal proteome stability across the tree of life. *Nat Methods* **17**, 495–503 (2020).
21. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
22. Gholizadeh, E. *et al.* Targeting AML Resistance with LY3009120-Sapanisertib and Ruxolitinib-Ulixertinib Combinations Demonstrate Superior Efficacy in*FLT3*,*TP53*,and*MUC4*mutations. *bioRxiv* (2024) doi:10.1101/2024.12.31.630711.

23. Mirzaie, M. *et al.* Designing patient-oriented combination therapies for acute myeloid leukemia based on efficacy/toxicity integration and bipartite network modeling. *Oncogenesis* **13**, 11 (2024).