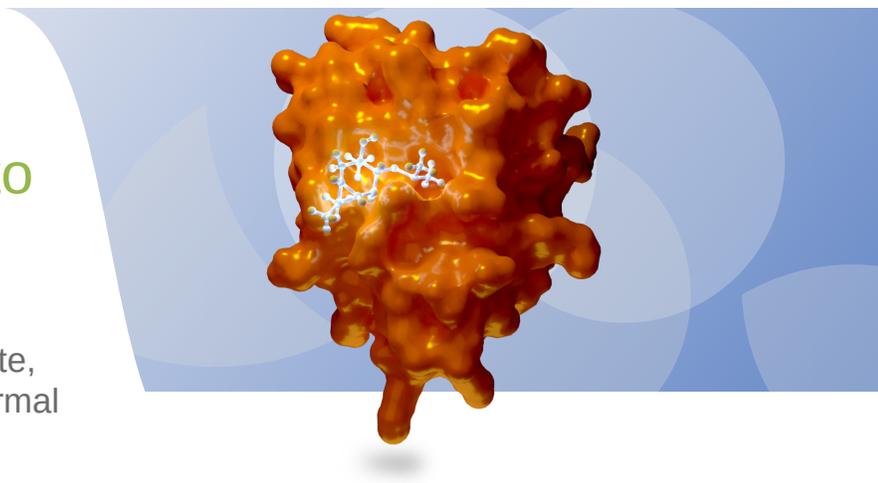


Proteome Screening with MatchMaker achieves performance comparable to experimental techniques

MatchMaker powered Proteome Screening identifies more known binders of methotrexate, dasatinib and panobinostat compared to thermal proteome profiling (TPP)



OPPORTUNITY

In silico methods capable of matching the performance of experimental approaches in drug target deconvolution can accelerate the therapeutic development process.

TECHNOLOGY

MatchMaker
Ligand Express Proteome Screening

SOLUTION

We compared MM Proteome Screening results with those from TPP. We show that MM exceeds the performance of TPP in identifying high confidence targets of methotrexate, dasatinib and panobinostat.

INTRODUCTION

Both *in silico* and experimental approaches have been proposed to elucidate a drug's polypharmacology. Screening drugs against many proteins permits the identification of unknown targets, informs the selection of candidate molecules for drug development, enables repurposing of approved drugs for new indications, and improves safety by identifying potentially harmful off-target interactions. Thermal proteome profiling is a recent high throughput experimental technique for detecting drug target engagement, which combines cellular thermal shift assay (CETSA) with quantitative mass spectrometry¹. TPP relies on altered protein thermostability upon ligand binding, applies to any class of target or compound, and does not require any modification to the compound. Other examples of chemical proteomics methods for target deconvolution include activity-based protein profiling, photoaffinity labeling and affinity purification, all of which require potentially activity-altering compound modification. Herein we compare MatchMaker powered Proteome Screening (MMPS) to TPP regarding their ability to identify the known protein drug targets of methotrexate, dasatinib, and panobinostat (**Figure 1**).

METHODOLOGY

We selected methotrexate, dasatinib, and panobinostat for this study because of readily available TPP data from human cell lines^{1,3,4}, and because of sufficient high quality experimental data available to use as a reference for ground truth. For reference, we use the comprehensive protein-chemical interaction database STITCH which integrates experimentally validated interactions (e.g. binding experiments evaluating K_i , IC_{50} , or K_D), manually curated datasets, and automated text mining. We limit our reference list to proteins with high confidence interaction and binding scores (>0.7), which excludes solely text mining

results. To estimate the performance of MMPS when applied to investigational compounds and Novel Chemical Entities (NCE), we retrained its deep learning model for the purpose of this benchmark while excluding bioassay and structural data for all compounds similar to the three test drugs. Specifically, we apply a cutoff of 0.5 Tanimoto Similarity (TS) or greater, using Morgan Fingerprints (Radius 3), to attenuate over-estimation of model quality due to compound series bias⁵. Recently reported Drug Target Interaction (DCI) predictions fail to exclude similar molecules between testing and training sets⁶ or from database searches⁷, and therefore reported performances would not be representative of NCEs, which are critical for pharmaceutical drug development. We then compared the lists of proteins identified by MMPS and TPP against the reference list of high confidence targets for each molecule. A protein was counted as identified by MMPS if it appeared within the top 25 of the 8700 proteins ranked. When multiple TPP screens were available for different conditions (e.g. intact cells versus cell extracts), we analyzed the results for each individual condition separately, in addition to the cumulative results across all conditions.

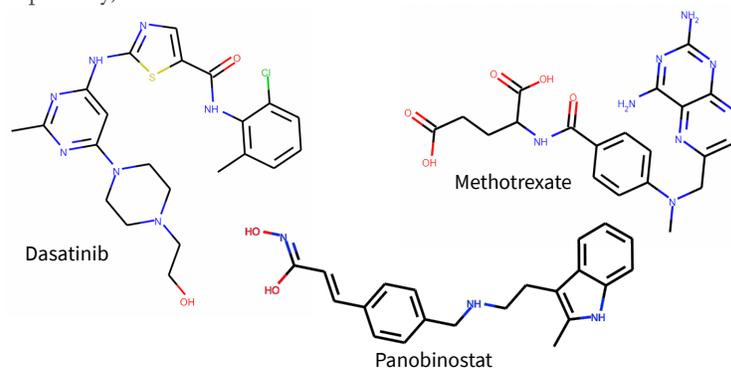


Figure 1. 2D chemical structures of dasatinib, methotrexate, and panobinostat.

RESULTS

Methotrexate is an antimetabolite multi-targeted drug that inhibits the synthesis of pyrimidines and purines through its inhibition of dihydrofolate dehydrogenase (DHFR). Out of the 14 high confidence binders of Methotrexate, two proteins (DHFR and TYMS) were identified by both TPP³ and MMPS, with MMPS identifying an additional two (FOLR2 and DHFR2) known binders (Figure 2A), highlighting the value that MMPS has in elucidating a drug's polypharmacology. Significantly, DHFR2 was found even though there is no crystal structure available for it in the PDB, demonstrating the flexibility of the MatchMaker engine.

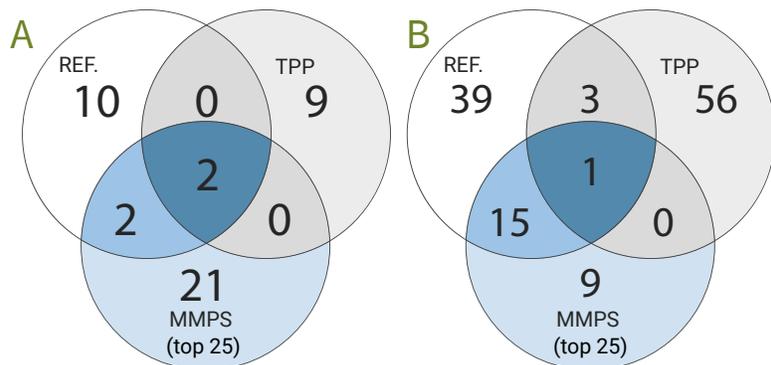


Figure 2. (A) The number of proteins interacting with methotrexate identified by either thermal proteome profiling (TPP) or MMPS and their overlap with the reference (REF.) interactors. All three identified dihydrofolate reductase (DHFR) and thymidylate synthase (TYMS) as binders. (B) Comparison of results generated for Dasatinib. All three identified Tyrosine-protein kinase Yes (YES) as a binder.

Dasatinib is a well-studied kinase inhibitor targeting BCR-ABL and SRC. It is used in the treatment of chronic myeloid leukemia (CML) and acute lymphoblastic leukemia (ALL). For dasatinib, out of the 58 high confidence reference interactors, a total of four overlapped with those identified cumulatively through TPP¹ (Figure 2B), spanning four individual TPP experimental conditions. MMPS identified 16, only one of which were also identified by TPP (YES), highlighting the complementary nature of the two approaches and greatly improving upon the false negative limitation of TPP. Extending the MMPS list to the top 100 predicted proteins for dasatinib captures 28 out of the 58 reference protein interactors. Notably, four of the reference proteins (SRMS, FRK, BLK and FGR) identified within the top 60 by MMPS were identified even though there is no structure available in the PDB.

Panobinostat is a pan-histone deacetylase (HDAC) inhibitor used in the treatment of multiple myeloma. Out of the 11 reference protein interactors of panobinostat, four were identified cumulatively through two separate TPP conditions. MMPS also identified four, with two proteins identified by both approaches (HDAC2 and HDAC10) (Figure 3), again highlighting the complementarity of the two approaches. Remarkably, HDAC10 was identified through MMPS even though there is no structure available for it in the PDB.

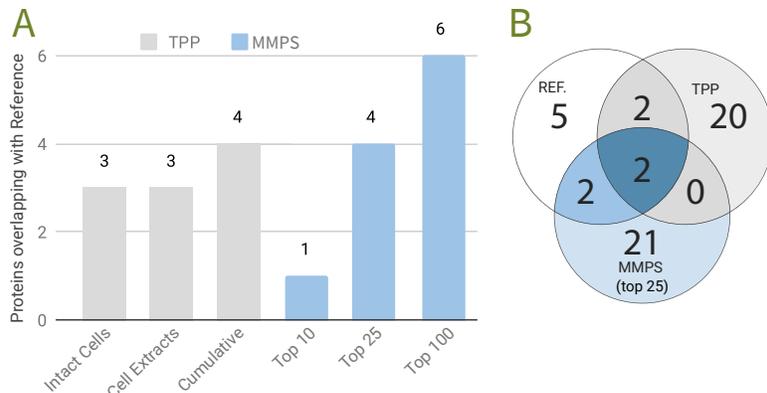


Figure 3. (A) The number of proteins interacting with panobinostat identified by either TPP or MMPS which overlap with the reference set. Extending MMPS to the top 100 predictions results in the identification of 6 out of the 11 high confidence reference interactors. (B) The number of proteins interacting with panobinostat identified by either TPP or MMPS which overlap with the reference set. All three methods identified Histone deacetylase 2 (HDAC2) and Histone deacetylase 10 (HDAC10) as binders.

SUMMARY

Protein drug interactions predicted independently by MMPS and by TPP were compared to high confidence protein drug interactions listed in the STITCH database for methotrexate, dasatinib, and panobinostat. While the false positive rate of TPP is low, we find that the method is susceptible to false negatives, consistent with previous observation⁸. MMPS identified several additional true positives compared to TPP, demonstrating the complementary nature of the two methods, but also the viability of MMPS as an *in silico* only approach. We cannot draw firm conclusions about the false positive rate of either method, because any positives not found in the reference list may well represent true, previously uncharacterized interactions. For both methotrexate and dasatinib, MMPS outperformed TPP in identifying high confidence protein targets, supporting the exciting possibility that using exclusively an *in silico* technology can yield substantial improvements over experimental methods in target deconvolution as well as the assessment of therapeutic safety and efficacy of new molecules.

RESOURCES

1. Savitski MM, *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science*. 346, 1255784 (2014)
2. Szklarczyk D, *et al.* STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* 44, D380-4 (2016)
3. Huber KV, *et al.* Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. *Nat Methods*. 12, 1055-7 (2015).
4. Becher I, *et al.* Thermal profiling reveals phenylalanine hydroxylase as an off-target of panobinostat. *Nat Chem Biol*. 12(11), 908-910 (2016)
5. Mayr, A, *et al.* Large-Scale Comparison of Machine Learning Methods for Drug Target Prediction on ChEMBL. *Chem. Sci.* 9, 5441-5451. (2018)
6. Hie, B., *et al.* Realizing private and practical pharmacological collaboration. *Science*. 350, 347-350 (2018)
7. Zhou, H., *et al.* FINDSITE comb2.0 : A New Approach for Virtual Ligand Screening of Proteins and Virtual Target Screening of Biomolecules. *J. Chem. Inf. Model.*, acs.jcim.8b00309. (2018)
8. Drug Selectivity: An Evolving Concept in Medicinal Chemistry., 2017, John Wiley & Sons., Handler N. *et al.*

CYCLICA INC.

18 King St East, Suite 801
Toronto, Ontario, M5C 1C4, Canada
1-416-304-9201