RepurposeVS: A Drug Repurposing-Focused Computational Method for Accurate Drug-Target Signature Predictions

Naiem T. Issa[1], **Oakland J. Peters**[1], **Stephen W. Byers**[1,2], and **Sivanesan Dakshanamurthy**[*,1,2]

[1]Department of Oncology, Georgetown Lombardi Cancer Center, Washington, D.C. 20057, USA

[2]Department of Biochemistry, Molecular and Cellular Biology, Georgetown University Medical Center, Washington, D.C. 20057, USA

Abstract

We describe here RepurposeVS for the reliable prediction of drug-target signatures using X-ray protein crystal structures. RepurposeVS is a virtual screening method that incorporates docking, drug-centric and protein-centric 2D/3D fingerprints with a rigorous mathematical normalization procedure to account for the variability in units and provide high-resolution contextual information for drug-target binding. Validity was confirmed by the following: (1) providing the greatest enrichment of known drug binders for multiple protein targets in virtual screening experiments, (2) determining that similarly shaped protein target pockets are predicted to bind drugs of similar 3D shapes when RepurposeVS is applied to 2,335 human protein targets, and (3) determining true biological associations *in vitro* for mebendazole (MBZ) across many predicted kinase targets for potential cancer repurposing. Since RepurposeVS is a drug repurposing-focused method, benchmarking was conducted on a set of 3,671 FDA approved and experimental drugs rather than the Database of Useful Decoys (DUDE) so as to streamline downstream repurposing experiments. We further apply RepurposeVS to explore the overall potential drug repurposing space for currently approved drugs. RepurposeVS is not computationally intensive and increases performance accuracy, thus serving as an efficient and powerful *in silico* tool to predict drug-target associations in drug repurposing.

Keywords

Cancer; drug; interaction; mebendazole; repositioning; repurposing; virtual screening

[*]Address correspondence to this author at the Department of Oncology, Georgetown University, Washington D.C. 20057, USA; Tel: + +1-202-687-2347; sd233@georgetown.edu.

1. INTRODUCTION

Drug repurposing- the process of utilizing drugs approved for one indication for another- is an efficient method for bolstering the pharmaceutical pipeline [1]. Given that approved drugs have known well-tolerated toxicity profiles, they can, therefore, be streamlined back into the development pipeline directly at phase II. Despite some successes, drug repurposing remains a challenge for two main reasons: (1) validating druggable therapeutic target(s) associated with the disease, and (2) confidently establishing the repertoire of protein target interactions for the FDA approved drug set. This manuscript will focus on the latter aspect.

A variety of methods for establishing drug-target interactions are employed in both academia and industry. High-throughput screening (HTS) strategies are used for establishing interactions for large drug libraries against protein targets of interest [2]. These approaches, however, have multiple obstacles. These include the financial cost per assay run, development of appropriate screening assays, maintaining biochemical relevance of the target given the assay (i.e. target immobilization in 96-well plates may alter binding site properties), among others [3]. The amount of potential druggable disease-related targets is also exponentially increasing [4] along with the number of synthesizable drugs [5]. Creating the vast possible drug-target space of true interactions and further narrowing it to that of physiologic- and disease-relevance remains a great challenge.

Computer-aided methods allow for a substantial increase in efficiency in establishing drug-target interactions and are constantly becoming more accurate as the biophysical mechanisms behind molecular recognition become better understood [6]. Such methods are typically used in virtual screenings against a protein target of interest, where large drug libraries (>1,000,000 structures) are subjected to an algorithm that quantifies the drugs' "fit" into the binding site. The first few hundred or thousand drugs are then validated experimentally, and the potential drug-target space has been drastically reduced to that with the greatest biological plausibility.

Many efforts for computationally predicting drug-target interactions exist, spanning both chemo-centric [7, 8] and target-based methodologies [9, 10]. Chemo-centric approaches utilize physical and chemical information obtained from ligands. Some approaches relate receptors based on the chemical similarity [7] as well as shape similarity [8] between ligands. Large public databases that aid in extracting ligand-based data for informatics also exist [9]. Target-based approaches, on the other hand, rely on docking [10–13] or binding site similarity [14]. Docking has driven some successful drug repurposing attempts [15–19], but scoring functions are generally considered inaccurate in calculating free energies of binding due to difficulty in predicting bioactive poses and variable contributions of weak interactions [20]. Alternatively, binding site comparison methods [21, 22] are implemented under the premise that similar binding sites should bind similar molecules. The use of binding site similarities has been successful in identifying novel targets for known drugs [23, 24] under the assumption that drugs interact with proteins containing similar binding sites [25, 26].

While chemo-centric and target-based methods have their own strengths and limitations, few computational methods attempt to combine ligand- and protein-based approaches [27, 28]. In this work, we present RepurposeVS, a comprehensive method for predicting FDA approved and experimental drug-protein target interactions through computationally efficient virtual screenings. RepurposeVS combines high-throughput docking with quantified shape, atom pair, and other descriptor similarity information of query drugs to reference experimentally derived crystal structure complexes. Furthermore, the utilized normalization procedure provides biological context of binding and allows for cross-protein comparison of drug binding signatures instead of protein-specific rank-ordering of drugs. This enables a standardized prioritization of predicted drug-target signatures for the entire proteome cohort in a study and the future incorporation of new signatures when novel protein target structures become available.

To assess accuracy, RepurposeVS was compared to the GLIDE docking algorithm in virtual screening experiments for prioritization of known drug binders for multiple pharmaceutically relevant protein targets. As RepurposeVS is a drug repurposing-driven method, the drug set chosen for benchmarking includes 3,671 FDA approved and experimental drugs. This drug set is composed of diverse chemical structures and chemotypes, as well as streamlines the generation of drug repurposing hypotheses for later experimental testing. Although benchmarks for virtual screening methods typically utilize the Database of Decoys (DUD-E) [29], we are focused on drug repurposing and therefore the ability of RepurposeVS to enrich for actives from an approved drug set rather than a chemical set of closely related analogues that may or may not be clinically relevant. RepurposeVS provided the greatest enrichment for known active drugs and was then scaled up to predict drug-target signatures across 2,335 human protein targets. Cursory global validation across the entire protein target set was then achieved by recapitulating the phenomenon of similarly shaped protein pockets binding drugs of similar shape [30, 31]. Biological validation was further obtained for the anti-hookworm drug mebendazole *via* kinase binding assays, thus providing further evidence to its anti-cancer efficacy for repurposing. Finally, RepurposeVS was used to explore the entire potential drug repurposing space by devising a "repurposing potential score". With its high accuracy and ease of implementation, RepurposeVS is an efficient computational method for the accurate prediction of drug-protein target signatures to drive drug repurposing efforts forward.

2. MATERIALS AND METHODS

2.1. Drug and Protein Target Dataset

Drugs were obtained from the DrugBank [32], FDA [33] and BindingDB [34]. LigPrep [35] was used to prepare and minimize drug structures at neutral pH of 7.0. Human protein target crystal structures containing a reference drug in the binding pocket with X-ray resolution <2.5 angstrom were chosen from RCSB (www.rcsb.org). After processing, the dataset included 3,671 drugs and 2,335 protein target crystal structures. Known active drugs for the benchmark protein targets HSP90A (PDB: 4O05), CA4 (PDB: 3FW3), ALDR1 (PDB: 3RX3), ACE (PDB: 1O86), PPARG (PDB: 3VSO), ADRB2 (PDB: 3NYA), VEGFR2 (PDB:

2P2H), ESR1 (PDB: 3ERD), AR (PDB: 3L3Z), BACE1 (PDB: 3VF3), GR (PDB: 4P6X), and HMGCR (PDB: 1HWK) were obtained *via* DrugBank annotations.

2.2. RepurposeVS Procedure

The workflow for RepurposeVS, modeled after the "Train Match, Fit, Streamline" (TMFS) protocol [36], is outlined in Fig. (1). A 3D comprehensive conformer library was generated using ConfGen [37] for each drug. From this library, the conformer whose 3D shape was most similar to that of the reference ligand bioactive pose was chosen for all subsequent steps. GLIDE [38] docking was performed to obtain free energies of binding, QikProp [39] was used for generating ligand-based 2D descriptors, and 3D shape descriptors for drug and protein binding sites were generated using spherical harmonics expansion coefficients Java software package provided to us by the Thornton group [40]. Reference-occupied protein target pocket shapes were determined using protomol information from sc-PDB [41]. Atom Pair (AP) similarity normalized scores were calculated directly using Strike [42].

The RepurposeVS Z-score ranking equation for a query drug q against protein target p with reference drug r is as follows:

$$Z(q,p,r) = \omega_j Y(p,q) + \omega_k P(r,q) + \sum_{m=1}^{1} [\omega_m f_m(p,q) + \omega'_m f'_m(r,q)] + \sum_{n=1}^{N} X_n(r,q) + CS(OLIC)$$

(1)

Y represents the rigorously normalized docking score based on the method outlined in Section 2.2.1 below with weight $\omega_j = 4$. P represents the normalized AP similarity tanimoto coefficient (T_c) of a query drug q to the reference drug r along with its designated weight ($\omega_k = 4$). The first summation corresponds to the shape similarity metric composed of two functions: (1) $\omega_m f_m(p,q)$ $\omega_m f_m (\sigma_p, \sigma_1)$, where f_m is the shape function corresponding to a similarity quantification between pocket shape of the protein target p and the query drug q with weighting factor $\omega_m = 2$, and (2) $\omega'_m f'_m(r,q)$, where f'_m is a shape function corresponding to a similarity quantification between reference drug shape r and query drug shape q with weighting factor $\omega'_m = 2$. Shape similarities are represented as Euclidean distances between the spherical harmonics expansion coefficients, as described in [43]. The second summation term corresponds to the combined similarity of $N = 10$ query drug-based descriptors terms (X_n) to reference drug r. Normalized T_c scores were calculated for the following descriptors: (1) number of H-bond acceptors, (2) number of H-bond donors, (3) dipole, (4) electron affinity, (5) globularity, (6) molecular weight, (7) ClogP, (8) number of rotatable bonds, (9) solvent-accessible surface area, and (10) volume.

The *CS*(*OLIC*) term is a correction term called "optimal ligand interaction correction" (OLIC), an algorithm that obtains a better estimate of drug-protein interactions on the reference binding site by assuming that drugs will have similar experimental activity if their interaction involves similar binding site residues and makes similar interaction patterns to

the reference drug. The following equation was used to determine binding site energies for reference drug (2) and query drugs (3):

$$S(OLIC-r)_p = \sum_{n=1}^{NR} \omega_n E_{n,p} \qquad (2)$$

$$S(OLIC-q)_{q,p} = \sum_{n=1}^{NQ} \omega_n E_{n,q,p} \qquad (3)$$

The sums are over the number of contact points NR or NQ between protein p and its reference drug or query drugs, respectively. Contact points for drugs are described as those that overlap with established reference drug-protein contacts. Drugs that match or cover most of the interactions as that of the reference scored higher. Their corresponding energies are evaluated and compared with the energy of the reference drug. Energy of the test ligands scored higher if it is close or higher than the energy of the reference drug. The $E_{n,p}$ corresponds to energy for the nth contact point for the reference drug-protein complex (p,r). The $E_{n,q,p}$ term corresponds to energy for the nth contact point for the qth query drug and protein p. Weighting factors specific to each contact point used and are dependent on the particular drug-target complex.

The correction term *CS(OLIC)* has been determined as a difference between the two sums:

$$CS(OLIC) = S(OLIC-r)_p - S(OLIC-q)_{q,p} \qquad (4)$$

The additive combination of the aforementioned normalized terms with their respective weights results in the final RepurposeVS comprehensive Z-score (1) to rank drugs for a given target.

2.2.1. Rigorous Normalization Procedure of RepurposeVS Terms—RepurposeVS contains distinct parameters in (1) that are represented in different units, which correspondingly contain very different raw numeric ranges. For example, docking scores are expressed in kJ/mol where small changes in number correspond to large changes in the free energies of binding. Shape similarity terms are quantified by Euclidean distances and, therefore, function on an independent range of values that are incompatible with other terms in the equation. Consequently, to better allow RepurposeVS parameters to be compared and weighted intelligently, raw values for the docking score and shape similarity terms Y, f_m, and f'_m were normalized onto the $N(x): R \rightarrow (0,1)$ unit range using a sigmoid function to preserve order and provide symmetry. The normalization function is defined as follows:

$$N_\alpha(x) = 1 - |1 - S_\alpha(x)| \qquad (5)$$

where x is the raw parameter, $S(x)$ is a sigmoid function, and a is a tunable scalar coefficient chosen to maximize the information-preserving variance in the image of $N(x)$ (5). Since the range varied significantly between parameters, the coefficient a varied as well.

For the sigmoid function, the hyperbolic tangent function is chosen because it is well-behaved and computationally tractable, yielding (6). Since some RepurposeVS parameters required subtly different normalization properties. Hence, (6) was re-expressed for easier modification in special normalization cases. Expressing (6) in terms of the simpler logistic function L, shown in (7), yields the equivalent function in (8):

$$N_\alpha(x) = 1 - |\tanh(\frac{\alpha x}{2})| \quad (6)$$

$$L_\alpha(x) = \frac{1}{1 + e^{-\alpha x}} \quad (7)$$

$$N_\alpha(x) = 1 - 2 * |\frac{1}{2} - L_\alpha(x)| \quad (8)$$

To check the information preserving quality of this normalization, we formed histograms of both the un-normalized, or raw (Fig. 2A), and normalized population distributions (Fig. 2B) for the shape similarity parameter using $\alpha = 0.1$. A scatter plot of the un-normalized shape parameter *versus* the normalized shape parameter was also formed (Fig. 2C). Fig. (2A) shows that in this case normalization results in a good fit for a symmetric and centered (at 0.5) Gaussian distribution implying that the normalized data will be statistically well behaved. By comparing the un-normalized to normalized distributions, we can see that the input distribution was not significantly distorted by our normalization function $N(x)$. Fig. (2C) shows an approximately linear relationship between the majority of raw and normalized data point pairs, implying that the coefficient was a good choice for capturing the dynamic range of most of the dataset. The procedure was repeated for docking scores Y using $\alpha = 0.25$ (data not shown). This normalization procedure allows for RepurposeVS to better predict viable drug-protein signatures in an absolute manner, where relativistic knowledge of other drugs in an experimental cohort is not necessary to quantify and establish binding signatures. Thus, resulting Z-scores can be pooled across all protein target systems for global objective prioritization of drug-target predictions.

2.3. Drug Shape Deviation Score

To determine shape similarity for drugs shared between unique protein target pairs, the "Drug Shape Deviation Score" $(\overline{F})$ metric was created. For analysis, target pairs must have at least three drugs predicted in common (i.e. within top 40 ranking for each protein target *via* Z-score). A "permutation of differences" (9)–(12) approach was applied to arrive at a

score within the 0-1 unit range that reflects the average shape deviation of the predicted common drugs for a protein target pair. The process is as follows:

$$V = \{v_1, v_2, \ldots, v_n\} \quad (9)$$

where, for a given protein target pair, V is the set of common drugs,

$$F = \{f_k = |a_{k_1} - a_{k_2}| | a_k = [v_i, v_j] \in C_2(V)\} \quad (10)$$

$$|F| = \binom{n}{2} = \frac{(n-1)n}{n} \quad (11)$$

$$\overline{F} = \sum_{k=1}^{|F|} \frac{f_k}{|k|} \quad (12)$$

a_k is the Euclidean distance between a pair of common drugs, $C_2(V)$ is the set of all combinations of two elements from V without replacement to generate the number of difference values, F is the set of differences between the Euclidean distances *via* all possible permutations, $|F|$ is the number of elements within set F, and $\overline{F}$ is the average across all Euclidean distance values.

2.4. Kinase Binding Assay

Kinase assays were performed using Kinomescan, by Discoverx, CA, USA and Caliper LabChip 3000 by Caliper Life sciences, USA as described previously [36]. The determination of MBZ thermodynamic binding affinities (K_d) to kinase targets predicted by RepurposeVS was performed by using active site-directed competition binding [44]. Kinase-tagged T7 phage strains were grown in parallel in 24-well blocks in an E. coli host derived from the BL21 strain. E. coli bacteria were grown to log-phase and infected with T7 phage from a frozen stock (multiplicity of infection = 0.4) and incubated with shaking at 32°C until lysis (90–150 minutes). The lysates were centrifuged (6,000 × g) and filtered (0.2 μm sieves) to remove cell debris. The remaining kinases were produced in HEK-293 cells and subsequently tagged with DNA for qPCR detection. Streptavidin-coated magnetic beads were treated with control (biotinylated) for 30 minutes at room temperature to generate affinity resins for kinase assays. The liganded beads were blocked with excess biotin and washed with blocking buffer (SeaBlock (Pierce), 1% BSA, 0.05 % Tween 20, 1 mM DTT) to remove unbound ligand and to reduce non-specific phage binding. Binding reactions were assembled by combining kinases, control liganded affinity beads, and mebendazole in 1x binding buffer (20 % SeaBlock, 0.17x PBS, 0.05 % Tween 20, 6 mM DTT). Mebendazole was prepared as 40x stocks in 100% DMSO and directly diluted into the assay. All reactions were performed in polypropylene 384-well plates in a final volume of 0.04 ml. The assay

plates were incubated at room temperature with shaking for 1 hour and the affinity beads were washed with wash buffer (1× PBS, 0.05 % Tween 20). The beads were then re-suspended in elution buffer (1× PBS, 0.05 % Tween 20, 0.5 μM non-biotinylated affinity ligand) and incubated at room temperature with shaking for 30 minutes. The kinase concentration in the eluates was measured by qPCR. Drugs that bind the kinase active site and directly prevent kinase binding to the immobilized ligand will reduce the amount of kinase captured, whereas drugs that do not bind the kinase have no effect on the amount of kinase captured. The amount of kinase captured in test *versus* control samples were measured by using a quantitative, precise and ultra-sensitive qPCR method that detects the associated DNA label. Using (13), the primary screen binding interactions are reported as '% Ctrl' (Percent kinase remaining activity), where lower numbers indicate stronger hits.

$$\text{Percent Control (\%Ctrl)} = \frac{\text{Mebendazole signal} - \text{Positive control signal}}{\text{DMSO Negative control signal} - \text{Positive control signal}} \times 100$$

(13)

In a similar manner, binding constants (K_d) for mebendazole-kinase interactions are calculated by measuring the amount of kinase captured as a function of the mebendazole concentration in a dose response manner. An 11-point 3-fold serial dilution of each test compound was prepared in 100% DMSO at 100x final test concentration and subsequently diluted to 1x in the assay (final DMSO concentration = 1%). Most K_ds were determined using a starting concentration = 30,000 nM. If the initial K_d determined was < 0.5 nM (the lowest concentration tested), the measurement was repeated with a serial dilution starting at a lower starting concentration. Binding constants (K_d) were calculated with a standard dose-response curve (drug dose (x-axis) - qPCR signal (y-axis)) using the Hill equation in (14) with the Hill Slope set to −1.

$$\text{Response (Y)} = \text{Background} + \frac{\text{Signal(max)} - \text{Background}}{1 + \left(\frac{k_d}{\text{Drug Dose (X)}}\right)\text{Hill Slope}}$$

(14)

2.5. Repurposing Potential

Original drug class indications, obtained from DrugBank [32], were given a "Repurposing Potential Score" (*T*) based on the number of drugs studied for a given approved indication class and the number of unique RepurposeVS-predicted disease classes for that indication class. (15) represents the repurposing potential score (*T*) for a given disease class *i*:

$$T = \frac{d_i}{d_{neo}} + \frac{k_i}{k_{neo}}$$ (15)

where d_i and d_{neo} correspond the number of drugs approved for disease classes "*i*" and neoplasms, respectively, and k_i and k_{neo} correspond to the number of predicted new disease

classes excluding the original for disease classes "i" and neoplasms. All disease classes are normalized to the neoplastic disease class since it contained both the greatest number of drugs with unique indications and predicted new diseases classes. The Online Mendelian Inheritance in Man (OMIM) [45] and the Comparative Toxicogenomics Database (CTD) [46] were used to annotate disease classes for predicted drug-protein target interactions.

3. REPURPOSEVS PERFORMS SUPERIORLY TO GLIDE DOCKING IN PRIORITIZING KNOWN BINDERS FOR PROTEIN TARGETS IN VIRTUAL SCREENING

Virtual screenings were performed on a set of 12 pharmaceutically relevant protein targets to assess the accuracy of RepurposeVS. RepurposeVS performed superiorly to GLIDE, a docking algorithm found to be accurate in high-throughput virtual screenings [47], in enriching for known drug binders to a protein target over a set of 3,671 drugs (Fig. 3A, B). Using a paired, one-tailed student's t-test, RepurposeVS performed statistically significantly better than GLIDE (P<0.05). Receiver operating curves demonstrate that RepurposeVS increased accuracy the most for solvent-exposed binding pockets, such as VEGFR2 kinase domain and β_2-adrenergic G protein-coupled receptor, whereas minimal increase occurred for buried pockets such as the estrogen and androgen nuclear receptors (Fig. S1). This differential may be attributed to greater flexibility in binding pose in exposed sites, which are specifically reflected by the docking score and pocket shape terms. Altering the weights ω_k and ω_m for docking score and pocket shape, respectively, had no appreciable effect on performance (Fig. 3A). This suggests that the other parameters in compensate for the imprecision derived from the nature of exposed pockets and that RepurposeVS is a robust method applicable to diverse protein targets.

4. GLOBAL VALIDATION OF REPURPOSEVS USING SHAPE SIMILARITY

RepurposeVS was applied to a set of 2,335 human protein target crystal structures and globally validated using the concept of similarly shaped drugs binding to protein target sites of similar shape. Shape complementarity is a critical aspect of biomolecular recognition, though it may not explain all possible binding modes. Nonetheless, it has generally been noted that drugs that interact with protein binding sites of similar shapes tend to exhibit shape similarity to each other. We first determined that the notion of similarly shaped drugs bind to similar protein pockets is upheld using the reference co-crystallized molecules for the protein target set (Fig. 4). Similarity between protein target pockets was quantified using two metrics: (1) Euclidean distance of the space-filling protomol structure (Fig. 4A), and (2) root-mean-square deviation (RMSD) of binding site residues 6Å from the geometric center of the bound molecule (Fig. 4B). The former metric characterizes the binding site occupancy volume whereas the latter metric is a topological term reflective of the binding site Cα backbone. RMSD values were calculated using Maestro [48]. There exists a direct correlation between drug-drug shape Euclidean distances and protein-protein binding site shape Euclidean distances (Fig. 4A) and backbone RMSDs (Fig. 4B). This implies that for true biochemical associations, determined *via* crystal structures, similarly shaped molecules bind protein pockets of similar shape and topology. Using the "Drug Shape Deviation

Score", $\overline{F}$ (12), a similar trend was observed for drugs predicted by RepurposeVS (top 40 by Z-score) to bind the same protein targets (Fig. 4). Thus, RepurposeVS is a valid method for determining drug-target associations across a large and diverse protein target set *via* the pharmacological metric of similarly shaped drugs binding similarly shaped protein pockets.

5. *IN VITRO* BIOLOGICAL VALIDATION OF REPURPOSEVS USING MEBENDAZOLE FOR CANCER DRUG REPURPOSING

To biologically confirm RepurposeVS predictions *in vitro*, we tested the binding of protein kinase target hits to mebendazole (MBZ) for cancer drug repurposing. MBZ was originally approved for its potent nanomolar inhibition of hookworm tubulin. It is thought that its cross-over effect on mammalian tubulin, though with 1000x less potency, is responsible for its anti-cancer efficacy *in vitro* [49]. Using kinase binding assays, nano- and micromolar inhibition of several predicted kinase targets of MBZ was confirmed (Table 1). MBZ appears to inhibit kinases found within two.main branches of the kinome phylogenetic tree, with nanomolar potency clustering on one branch and micromolar potency on the other (Fig. 5). However, intra-branch variability in potency is also obvserved. It is likely that the semi-promiscuous nature of MBZ (Fig. 5) towards kinases is a result of a small fragment that allows it to interact with the benzimidazole moiety acting as head group anchor connecting loop residues between the c-lobe and n-lobe. MBZ is also dually able to form water-mediated contacts or directly interact with ATP site cavity-forming residues in the absence of water molecules. Our predicted kinase hits and the activity data of MBZ indicate that its anti-cancer properties may be due to a synergistic inhibition of tubulin as well as kinase activity. Interestingly, for lung cancer, combined inhibition of microtubules and DYRK1B, a MBZ target (Table 1), is a more potent therapeutic strategy than microtubule inhibitors alone [50]. In this instance, a single drug such as MBZ, which has both properties, would be advantageous. RepurposeVS, thus, is able to reliably predict targets for MBZ that contribute to its repurposing for cancers.

6. DRUG REPURPOSING POTENTIAL

RepurposeVS was used to provide a cursory assessment the potential repurposing space for FDA approved drugs based on their drug classes (Fig. 6). We devised a repurposing potential score (*T*) in (15) for this purpose. Antineoplastic agents are shown to have the greatest repurposing potential with regards to the number of drugs and the diversity of newly predicted disease categories, with a total of 47 drugs and 8 disease categories. The nutritional-metabolic and neoplasm disease classes are also predicted to have the greatest number of drugs with the greatest number of unique original indications repurposed to them with 143 drugs/29 indications and 123 drugs/22 indications, respectively.

The overrepresentation of anti-neoplastic drugs is expected as tumor development is due to perturbations in a variety of cell processes that are likely shared with other diseases. Dysregulated kinase signaling, for example, is a ubiquitous pathogenic disease mechanism given the role of kinases in signal transduction. Thus, kinase inhibitors would be expected to potentially be useful in other diseases. In addition, some cancer drugs exhibit polypharmacology that simultaneously alter multiple cell processes. Alternatively, anti-

infection agents exhibit relatively low repurposing potential (Fig. 6). This emphasizes the selectivity of these agents towards non-human targets for efficacy and desired therapeutic indices [51]. Some of these drugs, however, exhibit modest repurposing potential. These include antibacterial agents, possibly attributed to structural similarity between bacterial motifs and human proteins [52]. Antipsychotic agents and other psychiatry-approved drugs also are predicted to have modest repurposing potential. These drugs typically exhibit polypharmacology through GPCR-mediated interactions [53], and some are being repurposed for cancer therapy [54]. The outcomes of the potential drug repurposing space are in pharmacological and clinical agreement with the known properties of the mentioned drugs, further confirming the ability of RepurposeVS to empirically predict drug-target signatures for higher-order pharmacologic assessment.

CONCLUSION

RepurposeVS is a combined drug-centric and protein-centric computational method for formulating drug-target signature predictions in drug repurposing. Validity was confirmed through benchmark virtual screenings using 12 protein targets of pharmaceutical interest to better enrich for their respective known approved drugs over GLIDE docking. RepurposeVS was also validated by recapitulating that drugs of similar shapes were predicted to bind similarly shaped protein pockets when defining pocket shapes through drug occupancy, and also by confirming predicted kinase hits of mebendazole *via* kinase binding assays. Finally, RepurposeVS was used to quantify "repurposing potential scores" for drugs categorized by disease indication and showed that anti-infection compounds had the least repurposing potential whereas anti-neoplastic drugs had the greatest. One limitation, however, is that diverse binding modes and protein flexibility are not accounted for. However, RepurposeVS aims only to reestablish the experimental binding states obtained from crystallography so as to decrease false positive and false negative outcomes in virtual screenings. Overall, we believe RepurposeVS to be an efficient computational method to aid drug repurposing endeavors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Biography

S. Dakshanamurthy

ABBREVIATIONS

AP	Atom-pair
FDA	Food and Drug Administration
HTS	High-throughput screening
MBZ	Mebendazole
Tc	Tanimoto coefficient
VEGFR2	Vascular endothelial growth factor receptor

References

1. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004; 3:673–683. [PubMed: 15286734]

2. Fox S, Farr-Jones S, Yund MA. High-throughput screening for drug discovery: continually transitioning into new technologies. J Biomol Screen. 1999; 4:183–186. [PubMed: 10838437]

3. Beibette J. Gaining confidence in high-throughput screening. Proc Natl Acad Sci U S A. 2012; 109:649–650. [PubMed: 22308304]

4. Griffith M, Griffith O, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA, Subramanian J, Govindan R, Kumar RD, Bose R, Ding L, Walker JR, Larson DE, Dooling DJ, Smith SM, Ley TJ, Mardis ER, Wilson RK. DGIdb: mining the druggable genome. Nat Methods. 2013; 10:1209–1210. [PubMed: 24122041]

5. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. J Chem Inf Model. 2012; 52:1757–1768. [PubMed: 22587354]

6. Bajorath J. Integration of virtual and high-throughput screening. Nat Rev Drug Discov. 2002; 1:882–894. [PubMed: 12415248]

7. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein Pharmacology by ligand chemistry. Nat Biotechnol. 25:197–206.

8. Warner WA, Sanchez R, Dawoodian A, Li E, Momand J. Identification of FDA-approved drugs that computationally bind to MDM2. Chem Biol Drug Des. 2012; 80:631–637. [PubMed: 22703617]

9. Bolton EE, Chen J, Kim S, Han L, He S, Shi W, Simonyan V, Sun Y, Thiessen PA, Wang J, Yu B, Zhang J, Bryant SH. PubChem3D: a new resource for scientists. J Cheminform. 2011; 3:32. [PubMed: 21933373]

10. Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecules. Proteins. 2001; 43:217–226. [PubMed: 11276090]

11. Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, Luo X, Zhu W, Chen K, Shen J, Wang X, Jiang H. TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res. 2006; 34:W219–W224. [PubMed: 16844997]

12. Paul N, Kellenberger E, Bret G, Muller P, Rognan D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. Proteins. 2004; 54:671–680. [PubMed: 14997563]

13. Kellenberger E, Foata N, Rognan D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. J Chem Inf Model. 2008; 48:1014–1025. [PubMed: 18412328]

14. Kellenberger E, Schalon C, Rognan D. How to measure the similarity between protein ligand-binding sites? Curr Comput Aided Drug Des. 2008; 4:209–220.

15. Yang L, Chen J, He L. Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. PLoS Comput Biol. 2009; 5:e1000441. [PubMed: 19629158]

16. Zahler S, Tietze S, Totzke F, Kubbutat M, Meijer L, Vollmar AM, Apostolakis J. Inverse *in silico* screening for identification of kinase inhibitor targets. Chem Biol. 2007; 14:1207–1214. [PubMed: 18022559]

17. Tang L, Li MH, Cao P, Wang F, Chang WR, Bach S, Reinhardt J, Ferandin Y, Galons H, Wan Y, Gray N, Meijer L, Jiang T, Liang DC. Crystal structure of pyridoxal kinase in complex with roscovitine and derivatives. J Biol Chem. 2005; 280:31220–31229. [PubMed: 15985434]

18. Do QT, Renimel I, Andre P, Lugnier C, Muller CD, Bernanrd P. Reverse pharmacognosy: application of selnergy, a new tool for lead discovery. The example of epsilon-viniferin. Curr Drug Discov Technol. 2005; 2:161–167. [PubMed: 16472225]

19. Cai J, Han C, Hu T, Zhang J, Wu D, Wang F, Liu Y, Ding J, Chen K, Yue J, Shen X, Jiang H. Peptide deformylase is a potential target for anti-Helicobacter pylori drugs: reverse docking, enzymatic assay, and X-ray crystallography validation. Protein Sci. 2006; 15:2071–2081. [PubMed: 16882991]

20. Bohari MH, Sastry GN. FDA approved drugs complexed to their targets: evaluating pose prediction accuracy for docking protocols. J Mol Model. 2012; 18:4263–4274. [PubMed: 22562231]

21. Meslamani J, Rognan D, Kellenberger E. sc-PDB: a database for identifying variations and multiplicity of 'druggable' binding sites in proteins. Bioinformatics. 2011; 27:1324–1326. [PubMed: 21398668]

22. Haupt VJ, Schroeder M. Old friends in new guise: repositioning of known drugs with structural bioinformatics. Brief Bioinform. 2011; 12:312–326. [PubMed: 21441562]

23. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. Drug discovery using chemical systems biology: repositioning the safe medicine comtan to treat multi-drug and extensively drug resistant tuberculosis. PLoS Comput Biol. 2009; 5:e1000423. [PubMed: 19578428]

24. Defranchi E, Schalon C, Messa M, Onofri F, Benfenati F, Rognan D. Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. PLoS One. 2010; 5:e12214. [PubMed: 20808948]

25. Das S, Kokardekar A, Breneman CM. Rapid comparison of protein binding site surfaces with property encoded shape distributions. J Chem Inf Model. 2009; 49:2863–2872. [PubMed: 19919089]

26. Kahraman A, Morris RJ, Laskowski RA, Favia AD, Thornton JM. On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. Proteins. 2010; 78:1120–1136. [PubMed: 19927322]

27. Muegge I. Synergies of Virtual Screening Approaches. Mini Rev Med Chem. 2008; 8:927–933. [PubMed: 18691150]

28. Broccatelli F, Brown N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. J Chem Inf Model. 2014; 6:1634–1641.

29. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. J Med Chem. 2006; 49:6789–6801. [PubMed: 17154509]

30. Das S, Kokardekar A, Breneman CM. Rapid comparison of protein binding site surfaces with property encoded shape distributions. J Chem Inf Model. 2009; 49:2863–2872. [PubMed: 19919089]

31. Haupt VJ, Daminelli S, Schroeder M. Drug promiscuity in PDB: protein binding site similarity is key. PLoS One. 2013; 8:e65894. [PubMed: 23805191]

32. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 2011; 39:D1035–D1041. [PubMed: 21059682]

33. U.S. Food and Drug Administration. www.FDA.gov (Accessed **2012**)

34. Liu T, Lin Y, Wen X, Jorrisen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. 2007; 35:D198–D201. [PubMed: 17145705]

35. Schrödinger Release 2013-3: LigPrep, version 2.8. Schrödinger, LLC; New York, NY: 2013.

36. Dakshanamurthy S, Issa NT, Assefnia S, Seshasayee A, Peters OJ, Madhvan S, Uren A, Brown ML, Byers SW. Predicting new indications for approved drugs using a proteochemometric method. J Med Chem. 2012; 55:6832–6848. [PubMed: 22780961]

37. Small-Molecule Drug Discovery Suite 2013-3: Confgen, version 2.6. Schrödinger, LLC; New York, NY: 2013.

38. Small-Molecule Drug Discovery Suite 2013-3: Glide, version 6.1. Schrödinger, LLC; New York, NY: 2013.

39. Small-Molecule Drug Discovery Suite 2013-3: QikProp, version 3.8. Schrödinger, LLC; New York, NY: 2013.

40. Kahraman A, Morris R, Laskowski R, Thornton J. Shape variation in protein binding pockets and their ligands. J Mol Biol. 2009; 368:283–301.

41. sc-PDB Home Page. http://bioinfo-pharma.u-strasbg.fr/scPDB/

42. Small-Molecule Drug Discovery Suite 2013-3: Strike, version 2.4. Schrödinger, LLC; New York, NY: 2013.

43. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. Bioinformatics. 2005; 21:2347–2355. [PubMed: 15728116]

44. Fabian MA, Biggs WH 3rd, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M, Ford JM, Galvin M, Gerlach JL, Grotzfield RM, Herrgard S, Insko DE, Insko MA, Lai AG, Lelias JM, Mehta SA, Milanov ZV, Velasco AM, Wodicka LM, Patel HK, Zarrinkar PP, Lockhart DJ. A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotechnol. 2005; 23:329–336. [PubMed: 15711537]

45. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University; Baltimore, MD: 2014. World Wide Web URL: http://omim.org/

46. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Rosenstein MC, Wiegers TC, Mattingly CJ. Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res. 2013; 41(D1):D1104–D1114. [PubMed: 23093600]

47. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A critical assessment of docking programs and scoring functions. J Med Chem. 2006; 49:5912–5931. [PubMed: 17004707]

48. Small-Molecule Drug Discovery Suite 2013-3: Maestro, version 9.6. Schrödinger, LLC; New York, NY: 2013.

49. Sasaki J, Ramesh R, Chada S, Gomyo Y, Roth JA, Mukhopadhyay T. The anthelmintic drug mebendazole induces mitotic arrest and apoptosis by depolymerizing tubulin in non-small cell lung cancer cells. Mol Cancer Ther. 2002; 1:1201–1209. [PubMed: 12479701]

50. Li L, Liu Y, Zhang Q, Zhou H, Zhang Y, Yan B. Comparison of cancer cell survival triggered by microtubule damage after turning Dyrk1B kinase on and off. ACS Chem Biol. 2014; 9:731–742. [PubMed: 24377315]

51. Pereira MP, Kelley SO. Maximizing the therapeutic window of an antimicrobial drug by imparting mitochondrial sequestration in human cells. J Am Chem Soc. 2011; 133:3260–3263. [PubMed: 21322645]

52. Trost B, Lucchese G, Stufano A, Bickis M, Kusalik A, Kanduc D. No human protein is exempt from bacterial motifs, not even one. Self Nonself. 2010; 4:328–334.

53. Roth BL, Sheffler DJl, Kroeze WK. Magic shotguns *versus* magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat Rev Drug Discov. 2004; 4:353–359.

54. Yeh CT, Wu AT, Chang PM, Chen KY, Yang CN, Yang SC, Ho CC, Chen CC, Kuo YL, Lee PY, Liu YW, Yen CC, Hsiao M, Lu PJ, Lai JM, Wang LS, Wu CH, Chiou JF, Yang PC, Huang CY. Trifluoperazine, an antipsychotic agent, inhibits cancer stem cell growth and overcomes drug resistance of lung cancer. Am J Respir Crit Care Med. 2012; 186:1180–1188. [PubMed: 23024022]
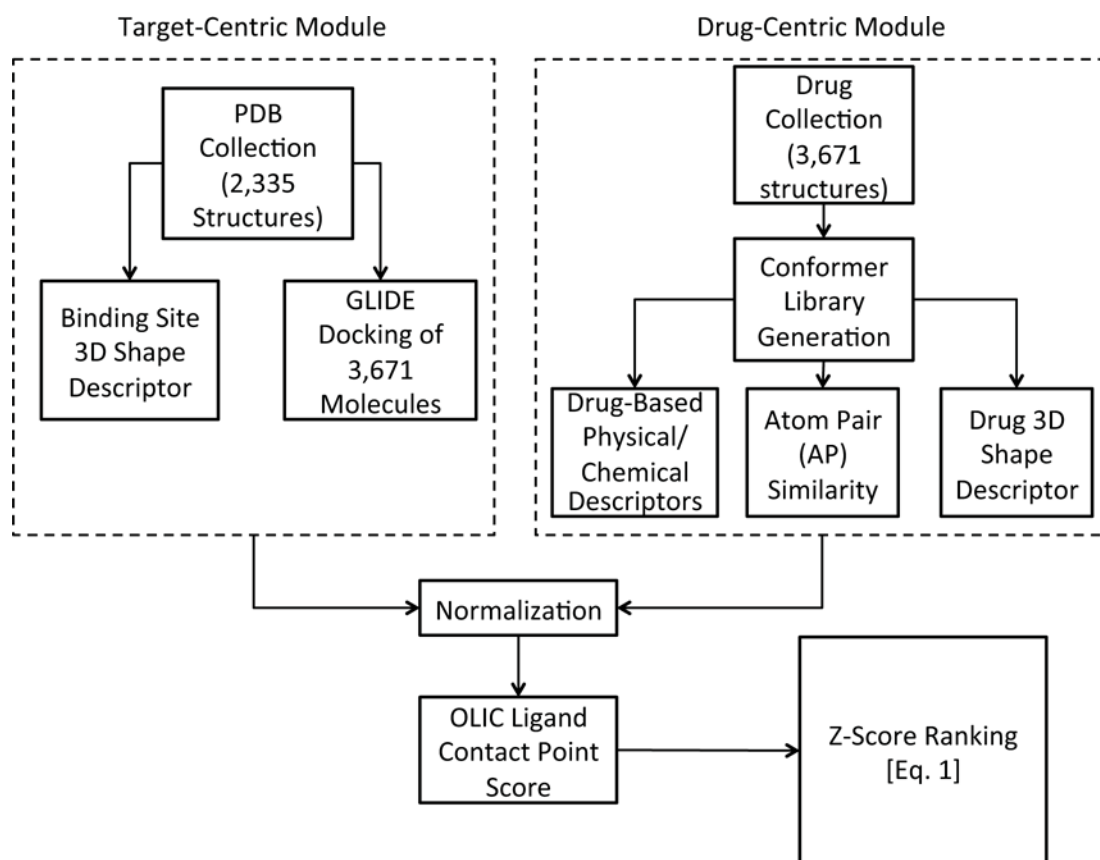
Fig. 1.
Workflow of RepurposeVS algorithm.
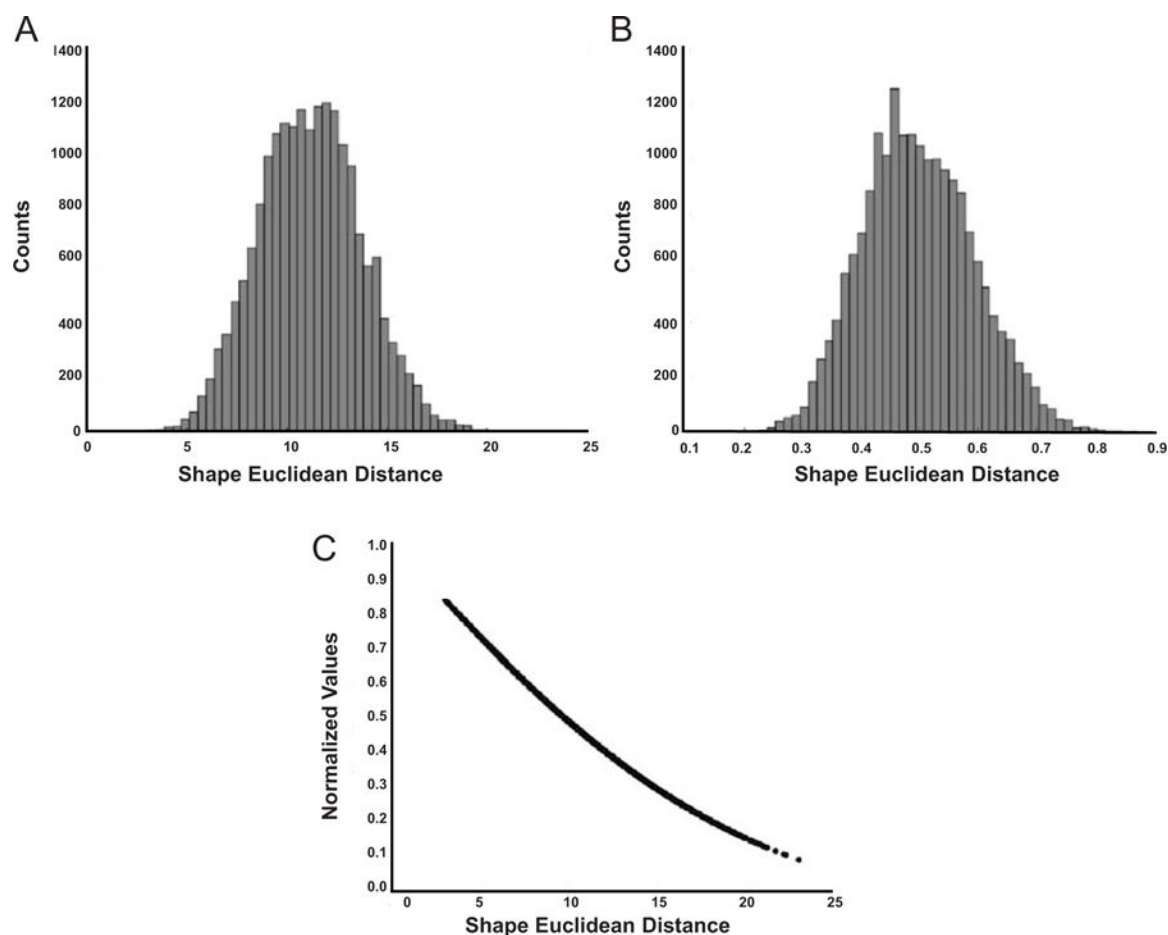
Fig. 2.
Normalization of RepurposeVS parameters. (**A**) Histogram of raw (non-normalized) scores for post-docking shape similarity Euclidean distance calculations of 2,207 unique drug-protein target pairs. The Y-axis shows counts of data points *versus* X-axis Euclidean distances using a bin-width of 0.5. (**B**) Histogram of normalized scores for the same 2,207 shape similarity calculations shown in (**A**). Normalization equation is shown in Eq. (14). The normalization preserves the Gaussian shape of the distribution, and centers the new distribution on the 0.5 mid-point of the 0-1 unit range. (**C**) Scatterplot showing the relationship between non-normalized shape similarity Euclidean distances (X-axis) and the resultant normalized values (Y-axis), for the data points shown in (**A**) and (**B**). The approximately linear relationship shown implies that the normalization does little to distort the population, although some bending is visible at the high-end (shape Euclidean distance values above 15).
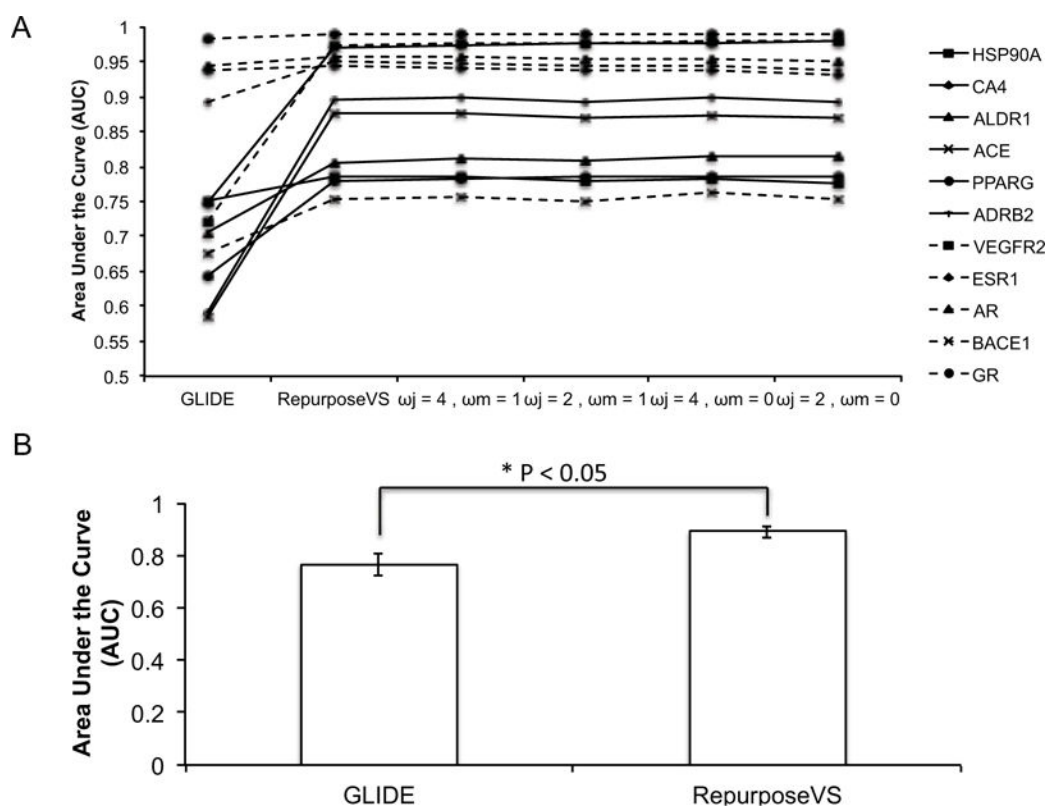
Fig. 3.
Areas under the curve (AUCs) for virtual screening of approved active drugs across 12 protein targets. (A) Outcomes of GLIDE docking and RepurposeVS in virtual screening experiments enriching for true active drugs for the noted protein targets. The remaining conditions reflect adjusted weights for the docking parameter (ωj) and protein shape parameter (ωm) in RepurposeVS (Eq. 1). (B) Average AUC across all 12 targets for each method.
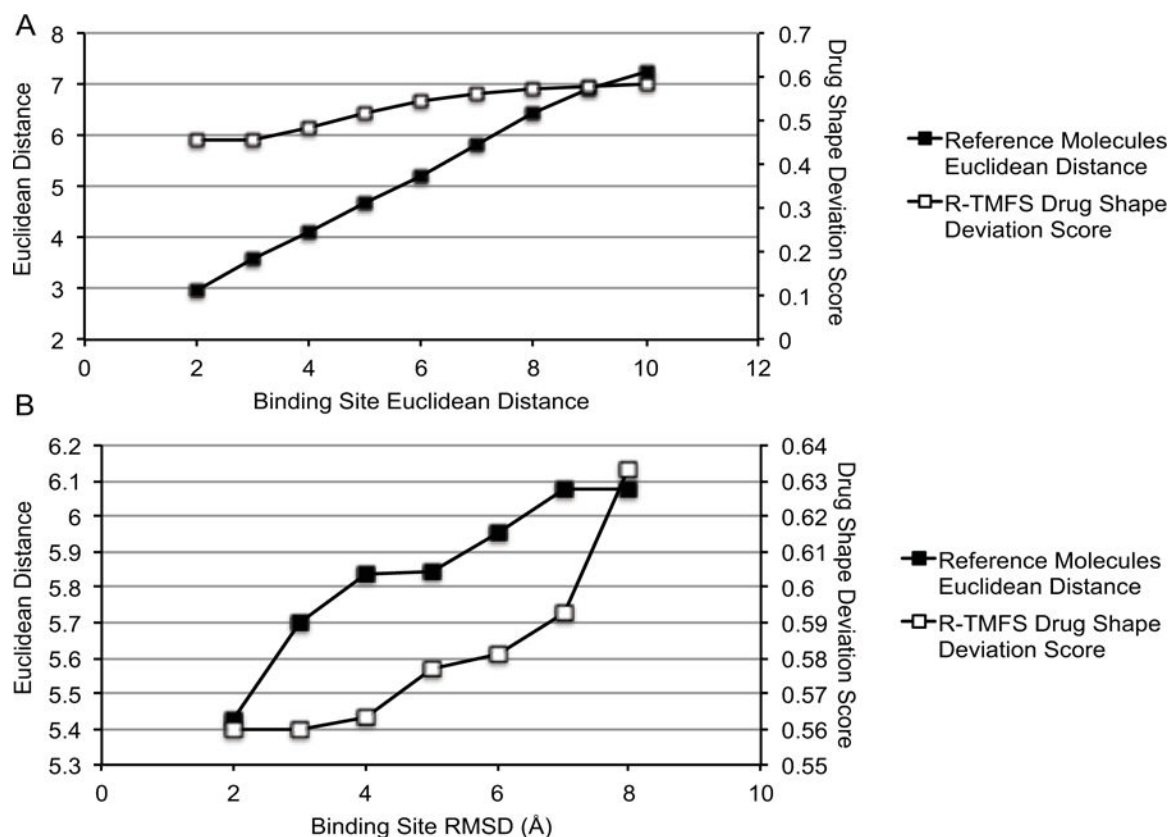
Fig. 4.

Trends in drug shape as a function of binding site shape and structural differences between protein target pairs. Line plots depicting shape Euclidean distances between co-crystallized reference molecules and normalized "Drug Shape Deviation Scores" ($\overline{F}$) against (A) binding pocket shape differences quantified by Euclidean distances and (B) backbone root-mean-squared deviation (RMSD) in angstroms. The data was binned into 1-unit groups with their means represented in the plot. Smaller Euclidean distances or RMSDs imply greater similarity.
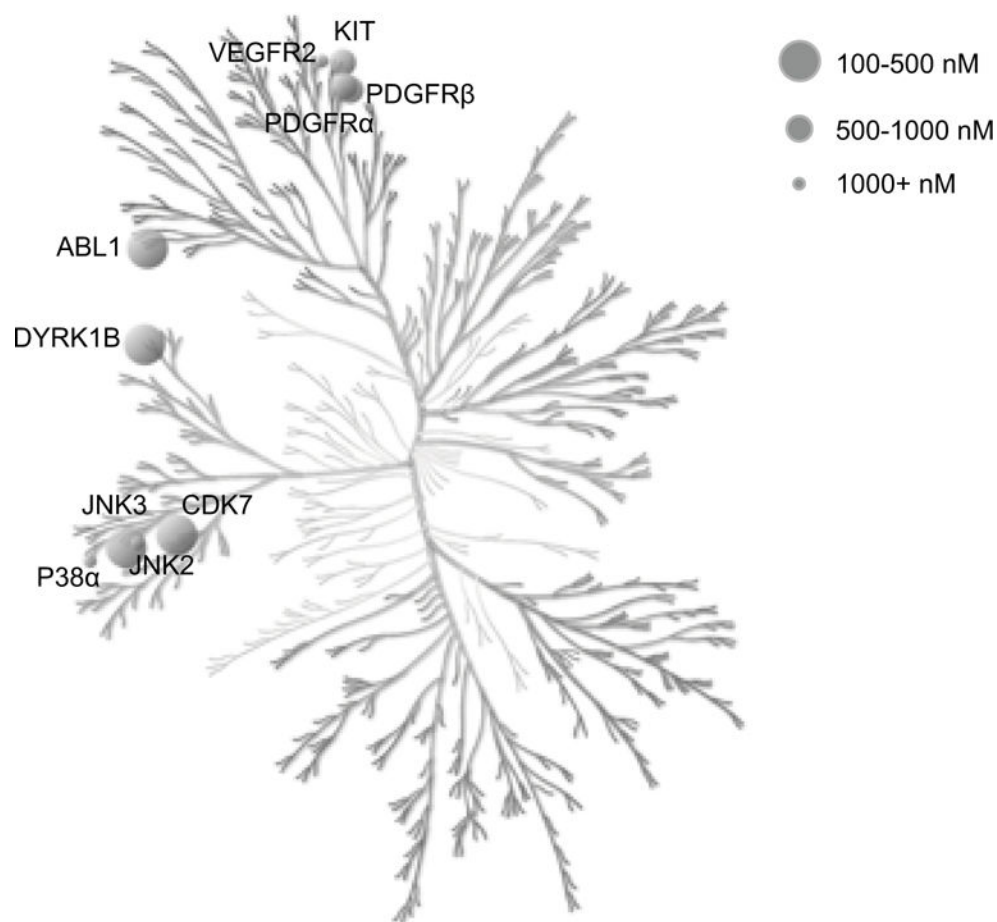
Fig. 5.
Validated mebendazole (MBZ) kinase targets predicted from RepurposeVS. Kinases for which binding affinities were determined are shown on the kinome phylogenetic tree.
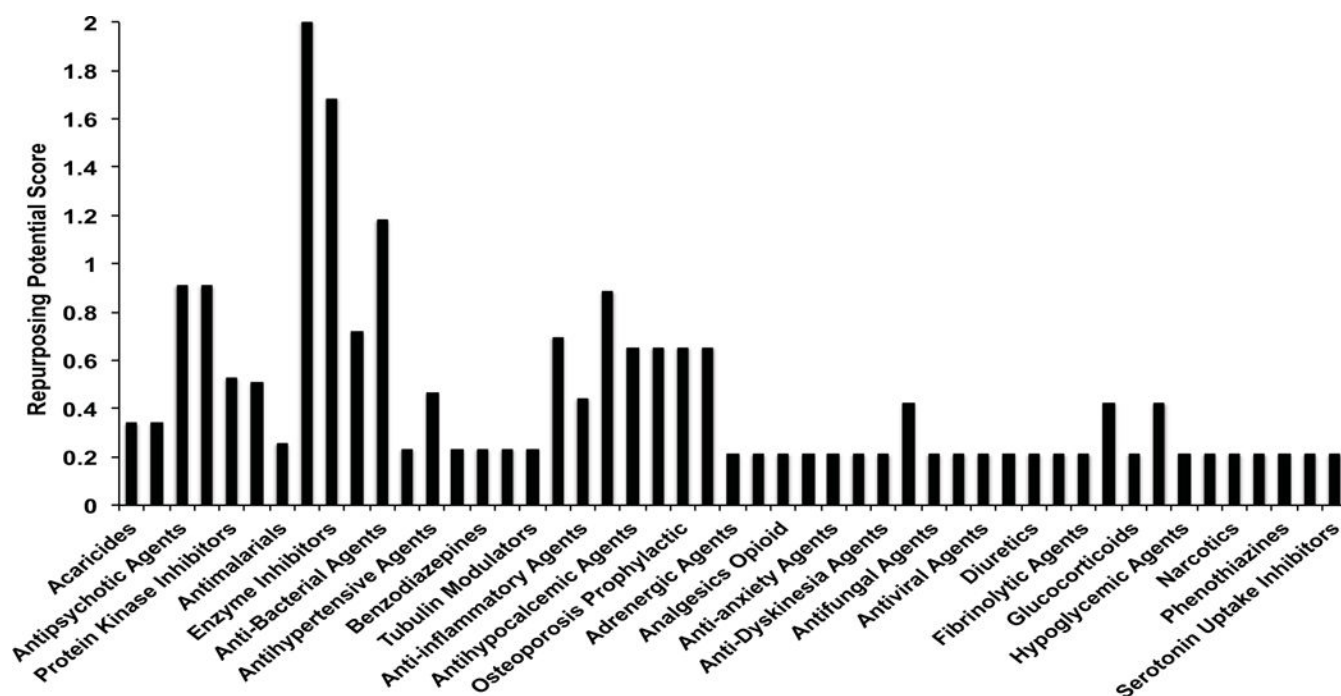
Fig. 6.
Histogram of predicted repurposing potential of approved drugs/indications to new disease classes. The "Repurposing Potential Score" is calculated using (15) (see Materials and Methods).

Table 1

Binding affinities of MBZ for predicted kinase hits.

Kinase Target	Percent Control at 10μM	Binding Affinity (K_d) in nM
ABL1(E255K)-phosphorylated	2.2	N/D
ABL1(T315I)-phosphorylated	3.2	N/D
ABL1-nonphosphorylated	2	N/D
ABL1-phosphorylated	0.9	120
CDK7	11	390
CSNK1D	36	N/D
DYRK1A	34	N/D
DYRK1B	5.6	340
GSK3B	35	N/D
JAK3	29	N/D
JNK1	14	N/D
JNK2	9.6	1090
JNK3	3	410
KIT (D816V)	7.4 (33)	750
MET	32	N/D
P38-alpha	17	1660
PDGFR-A	7.8	820
PDGFR-B	3.2	660
PIK3CG	18	N/D
SRC	34	N/D
ULK2	30	N/D
VEGFR-2	30	3600