



Review

Machine and deep learning approaches for cancer drug repurposing

Naiem T. Issa^a, Vasileios Stathias^b, Stephan Schürer^b, Sivanesan Dakshanamurthy^{c,*}^a Dr. Phillip Frost Department of Dermatology and Cutaneous Surgery, University of Miami School of Medicine, Miami, FL, USA^b Department of Molecular and Cellular Pharmacology, University of Miami School of Medicine, Miami, FL, USA^c Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA

ARTICLE INFO

Keywords:

Drug repurposing
Drug discovery
Machine learning
Deep learning
Artificial intelligence

ABSTRACT

Knowledge of the underpinnings of cancer initiation, progression and metastasis has increased exponentially in recent years. Advanced “omics” coupled with machine learning and artificial intelligence (deep learning) methods have helped elucidate targets and pathways critical to those processes that may be amenable to pharmacologic modulation. However, the current anti-cancer therapeutic armamentarium continues to lag behind. As the cost of developing a new drug remains prohibitively expensive, repurposing of existing approved and investigational drugs is sought after given known safety profiles and reduction in the cost barrier. Notably, successes in oncologic drug repurposing have been infrequent. Computational *in-silico* strategies have been developed to aid in modeling biological processes to find new disease-relevant targets and discovering novel drug-target and drug-phenotype associations. Machine and deep learning methods have especially enabled leaps in those successes. This review will discuss these methods as they pertain to cancer biology as well as immunomodulation for drug repurposing opportunities in oncologic diseases.

1. Introduction

In past years, medical oncology has witnessed an unprecedented explosion in the understanding of cancer pathophysiology and pathogenesis. With the advancement of next generation sequencing technologies such as single-cell RNA sequencing, we are better equipped to explore and model complex phenomena such as cancer heterogeneity, resistance and etiologies at a granular level. Moreover, the establishment and expansion of multi-institutional and large-scale biospecimen collection and bioinformatics initiatives, such as the Cancer Genome Atlas (TCGA), have allowed for the aggregation, curation and analysis of an unprecedented amount of patient-derived data that has led to the identification of novel therapeutic targets as well as the implication of well-known targets in new disease areas.

In spite of this extraordinary growth in the cancer biology field, progress in the area of drug discovery remains stagnant, still plagued with long development time to market and exorbitantly high costs despite the systematic implementation of high-throughput screening technologies. To date, bringing a drug to market takes about a decade

with research and development (R&D) costs reaching approximately US \$2.8 billion [1]. Candidate drugs may fail at many points along the drug development pipeline due to numerous reasons such as poor pharmacokinetics, toxicity or lack of clinical efficacy.

A promising solution to the considerable drug development challenges of novel compounds is the use of existing drugs for the treatment of new diseases. Approved drugs have undergone all phases of clinical trials in order to reach the market and thus have a known and accepted safety profile. If a new clinical indication for an approved drug is suggested, that drug may re-enter the clinical trial process at Phase II thus substantially reducing the R&D risk, time and cost [2].

In the past few years, drug repurposing research has benefited greatly by the systematic adoption of computational strategies spanning a large area of unique methodologies and approaches. Molecular modeling of therapeutic protein targets allows for the understanding of the structural biology as well as high throughput “virtual screenings” to identify novel drug candidates. Advances in machine learning and artificial intelligence (in particular deep learning) provide new insights into how drugs bind to targets and how their physicochemical properties

Abbreviations: ADE, adverse drug event; ADMET, absorption, distribution, metabolism, excretion, toxicity; CNN, convolutional neural network; DNN, deep neural network; EHR, electronic health record; EMT, epithelial-mesenchymal transition; ICI, immune checkpoint inhibitor; MDSC, myeloid-derived suppressor cells; NLP, natural language processing; PCM, proteochemometric; QSAR, quantitative structure-activity relationship; RF, Random Forest; SVM, Support Vector Machine; TCGA, The Cancer Genome Atlas; TIL, tumor immune cell infiltrates.

* Corresponding author.

E-mail address: sd233@georgetown.edu (S. Dakshanamurthy).

<https://doi.org/10.1016/j.semcancer.2019.12.011>

Received 18 July 2019; Received in revised form 31 October 2019; Accepted 15 December 2019

Available online 3 January 2020

1044-579X/© 2019 Elsevier Ltd. All rights reserved.

relate to phenotypic changes. In addition, the utilization of such methods also help identify novel anti-cancer targets from the large-scale cancer datasets already collected by multiple initiatives. As the amount of chemical and bioactivity data grows due to the adoption of high-throughput and multi-omics drug profiling assays, these methods are poised to make substantial contributions in cancer drug discovery. Furthermore, the increased accessibility of these public dataset collections further facilitates the potential of computational approaches. These approaches need not be limited in using experimental and biological datasets. The utilization of clinical datasets such as EHR (electronic health records) has also shown promising results. In this review, we discuss the current computational strategies used for drug repurposing in oncology with an emphasis on machine and deep learning.

2. Structure-based machine learning

Structure-based virtual screening methods have augmented the drug discovery process since the inception of *in-silico* molecular modeling (Fig. 1). In the most general sense, structure-based virtual screening aims to prioritize small molecules potentially targeting a protein target of interest, subsequently leading to the reduction of experimental costs and increase of successful outcomes. The typical approach includes the prediction of the binding affinity (or related affinity scores) between small molecules and the target of interest. This binding affinity/score is then used to rank-order compounds for prioritization, and the general rule is that the greater the predicted affinity the greater the likelihood of a true small molecule-target interaction.

Molecular docking is one of the most widely used structure-based techniques for predicting and prioritizing drug-target interactions. Briefly, the interactions between a drug and a target's binding sites are being assessed and quantified calculating the free energy of binding (ΔG), where lower or more negative ΔG values correspond to greater binding affinity. A big advantage of molecular docking methods is the interoperability of the predicted results as ΔG values are standardized and can be compared across millions of drugs that can be virtually screened. To calculate the ΔG , docking programs utilize scoring functions, which are mathematical formulas that incorporate physicochemical parameters of the drug and protein. Multiple docking programs exist that use a variety of scoring functions such as Glide [4] and AutoDock

[5]. The differences in scoring functions arise due to which parameters are used and how they are related to binding affinity. Overall, docking has proved to be a very effective drug repurposing tool for many diseases such as HIV [7], Dengue virus [8] and multiple cancer types [9]. One such recent example is the study from Hafeez et al. where they used docking approaches to reveal how ormeloxifene is able to modulate the epithelial to mesenchymal transition (EMT) process implicated in cancer development and progression [10]. They showed that ormeloxifene, an anti-contraceptive selective estrogen receptor modulator, to reduce prostate cancer growth through inhibition of beta-catenin signaling. Using docking, they revealed it could bind to beta-catenin and GSK3beta, thus explaining its unique effect on the EMT process, thus rendering it a promising drug repurposing pursuit for prostate cancer especially given its favorable safety profile.

3. Molecular docking using machine learning

While docking has potential for finding new targets for drugs, the major inherent limitation at this time is the accuracy of the scoring functions used. As previously stated, existing classical scoring functions are numerous and based on a presumptive functional relationship between binding affinity and the structure of drug-target complexes. This may lead to a considerably high false positive rate [11] and thus, utilizing new approaches, such as machine learning, are essential for improving the docking accuracy. By leveraging machine learning, scoring functions can be substantially improved in an unbiased way since drug-target relationships are inferred directly from the training data and are not influenced by presumptive relationships. Moreover, the predictive models, can be used to identify the key features of the datasets that contribute the most to the binding affinity and therefore provide a better understanding of the forces that drive the drug-target interactions.

The adoption and use of deep learning methods for molecular docking has been extremely successful mainly due to the accessibility to deep learning computational frameworks such as Tensorflow. Typically, these methods are applied to post-docking drug-target complexes in order to more accurately predict binding poses and drug activity. Such methods play a pivotal role in drug repurposing, as existing drugs can be virtually screened quickly across a large number of new clinically

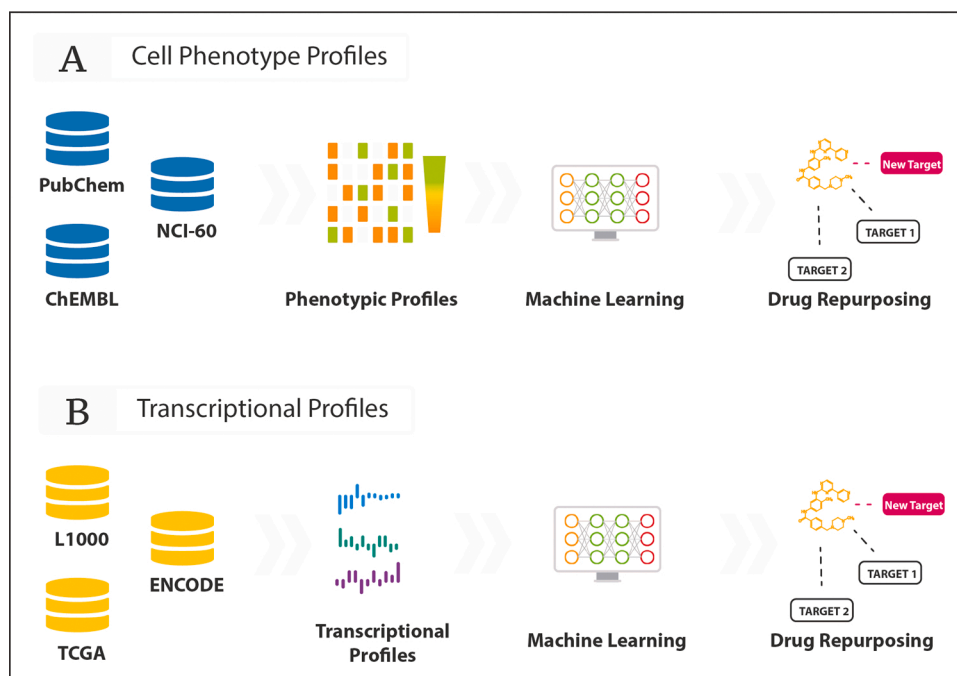


Fig. 1. Overview of structure-based machine learning strategies for identification of drug-target interactions. (A) Docking of small molecules into binding sites of three-dimensional protein target models. (B) Cheminformatics methods (i.e. QSAR) where chemical descriptors of small molecules are leveraged to create models relating descriptors to quantifiable target binding biological endpoints (i.e. binding vs non-binding, IC50, etc.). (C) Proteochemometrics similar to (B) but with additional descriptors derived from the protein target such as amino acid sequence.

relevant targets. In such a study, Wallach et al. utilized Convolved Neural Networks (CNN) to develop AtomNet [12]. AtomNet utilizes a three-dimensional (3D) grid representation placed on drug-target complex to extract structural features such as atom types and structural protein-ligand interaction fingerprints (SPLIFs), which are then unfolded into a one-dimension floating point vector. The training set was obtained from the scPDB database [13], which includes annotated 3D crystal structures of true drug-target interactions. AtomNet performed significantly better than commonly used docking algorithms such as Surflex-Dock [14] and DOCK [15] in prioritizing true active compounds from a set of decoys. Similarly, Ragoza et al. employed a CNN to 3D representations of drug-target interactions [16]. Their model was able to outperform AutoDock Vina [5] in discriminating between correct and incorrect binding poses as well as between active and inactive compounds. The aforementioned CNN-driven deep learning models were tuned to a dichotomous outcome (*i.e.* active versus inactive) and, therefore, are inherently limited in that they are unable to provide information about the strength of a drug-target interaction. More recently, Stepniewska-Dziubinska et al. developed Pafnucy, also a CNN based on 3D convolution, to estimate binding affinities of drug-target complexes. Pafnucy had outperformed all state-of-the-art functions when evaluated against the CASF-2013 “scoring power” benchmark [17]. Nguyen et al. were also successful in applying deep learning to enhance docking scoring functions for the prediction of binding poses, binding affinities and free energies of binding for a variety of cancer-related targets [18].

4. Ligand-based cheminformatics modeling using machine learning

A second approach for elucidating new drug-target interactions is that of ligand-based cheminformatics methods. In contrast to molecular docking methods, ligand-based cheminformatics methods do not require target protein-derived information. Rather, features generated from the chemical structure graph or any other suitable description of a drug are related to known target binding data (*i.e.* binding affinity) or phenotypic changes (*i.e.* high-throughput cancer cell viability assays) (Fig. 1). These features, also termed descriptors, can include physical, chemical, topological or other types of quantifiable information, typically derived from a drug's chemical structure. Other examples include atom types, electronic parameters, hydrophobicity, shape and many others. The collection of these values make up a chemical fingerprint for a drug and can be used for downstream analysis.

These structure-derived fingerprints are then related to target binding or phenotype data where the underlying assumption is that drugs with similar structures/properties exhibit similar bioactivity [19]. Similarity between two drugs, via their chemical fingerprints, is quantified by metrics such as the Tanimoto coefficient. Numerous drug-target associations have been elucidated using the chemical similarity technique [20]. While chemical similarity has been successful, there are instances where small chemical modifications (*i.e.* minor functional group changes) can result in large changes in activity, also known as “activity cliffs”, leading to false-positive associations [21].

In order to circumvent the above limitations, machine learning has been incorporated to better correlate drug features with bioactivity. This is often referred to as quantitative structure-activity relationship (QSAR). Generally speaking, machine learning is used either for categorical classification of drugs with respect to a target (*i.e.* inhibitor versus non-inhibitor) or predicting values along a continuous variable (*i.e.* binding affinity). Binding data used to produce machine learning models can be obtained from numerous sources including primary literature and online databases. These databases include BindingDB [22], PubChem [23], PharmGKB [24], ChEMBL [25], among others. Machine learning applied to these rich bioassay datasets has resulted in the successful identification of novel drug-target associations in multiple studies and has paved the way for a systematic drug repurposing

framework where features from existing drugs can be leveraged.

In one such study, Deshmukh et al. applied Support Vector Machine (SVM) and Random Forest (RF) machine learning algorithms to PubChem bioassay data to create classification models of inhibitors/non-inhibitors of flap endonuclease 1 (FEN1), an enzyme critical for DNA repair/replication and whose inhibition has been shown to increase cancer cell sensitivity to chemotherapy [26]. The SVM models outperformed the RF model and it was further used to screen a database of >50,000 compounds. The top five-ranked compounds were selected for *in-vitro* screening of which one (JFD00950) was identified as a novel FEN1 inhibitor with cytotoxic activity against the DLD-1 colon cancer cell line. In another study, Algamal et al. used adjusted adaptive least absolute shrinkage and selection operator (AALASSO) to predict the anti-cancer potency (IC₅₀) of imidazo[4,5-*b*]pyridine derivatives using 4885 chemical descriptors [27]. A three-dimensional (3D)-QSAR model of maslinic acid and its derivatives for their activity against the MCF-7 breast cancer cell line was developed by Alam and Khan using topological molecular features such as shape, electrostatic pattern and hydrophobic regions [28]. Using this model, they virtually screened the complete ZINC database [29] of millions of compounds and identified 593 potential hits. After applying ADMET filters, they found P-902 as a novel lead candidate with the potential to bind the NR3C1 glucocorticoid receptor using *in-silico* docking. Similarly, Taha et al. performed 3D-QSAR modeling on inhibitors of phosphoinositide 3-kinase gamma (PI3K γ) and screened the National Cancer Institute (NCI) list of compounds to identify 19 compounds with nanomolar inhibition of PI3K γ [30].

Ligand-based deep learning approaches have also been of benefit. In one such model, Allen et al. created a multi-task deep neural network (MTDNN) models to predict kinase activity of small molecules across the human kinome [31]. Briefly, comprehensive kinase bioactivity data was obtained from ChEMBL and the commercial database Kinase Knowledgebase KKB [32]. The total aggregated data included 668,920 measurements distributed across 342 targets and 317,228 unique compounds. In comparison to other machine learning methods, MTDNN had superior performance and was -+most generalizable. In another study, Rifaioğlu et al., leveraged two-dimensional structural compound representations as input in order to predict drug-target interactions using their convolutional neural network-driven method, DEEPscreen [33]. Apart from kinase bioactivity data, deep learning models such as MTDNN and DEEPscreen would enable large scale virtual screening of large compound databases for predictive activity profiling against kinases important for multiple cancer types and thus facilitate the quick and cost-effective repurposing of existing drugs.

In addition to inhibiting relevant protein cancer targets, small molecule modulation of microRNAs (miRNAs) is also of significant interest in cancer therapeutics. In general, MicroRNAs can alter oncogene function post-transcriptionally through their binding to the 3' non-translated region of transcripts [34] and therefore their inhibition could prove to be of great therapeutic benefit. Jamal et al. were the first to create a predictive model of small molecule modulators of miRNAs [35]. They integrated bioassay datasets found in PubChem and 179 molecular descriptors into machine learning algorithms and identified chemical scaffolds that may be critical for miRNA binding. Similarly, Wang et al. employed a random forest machine learning method using small molecule-miRNA binding data from the SM2miR database [36] and were able to predict the experimentally validated miRNA associations for commonly used small molecules 5-fluorouracil, 17 β -estradiol, and 5-aza-2'-deoxycytidine [37]. In another study, Qu et al. similarly created a machine learning framework called HeteSim that effectively predicts miRNA - small molecule associations [38]. A detailed overview of computational methods used for the accurate prediction of small molecule-miRNA binding pairs can be found in the review by Chen et al. [39].

5. Proteochemometrics modeling using machine learning

As a logical continuation of the QSAR methodology, the combination of features extracted from both the drug and the target protein has been particularly beneficial for developing predictive models of drug activity. These combinatorial approaches are known as proteochemometrics (PCM) and can uncover target features relevant to binding such as important amino acids. PCM has been used to find numerous novel inhibitors especially for adenosine receptors. Adenosine G protein-coupled receptors have always been a prominent target for cancer research as their upregulation/activation has been implicated in various tumor processes [40]. For example, activation of the adenosine A2A receptor suppresses regulatory T cell function and inhibits cytotoxic NK cell activity, thus attenuating anti-tumor immunity [41]. Van Westen et al. developed a PCM model from bioactivity data of all adenosine receptor subtypes (humans and rats) that was used to screen a database of >100,000 compounds from which six were experimentally confirmed to be high-affinity adenosine receptor binders [42]. Similarly, Cortes-Ciriano et al. used PCM to predict inhibitor potency against cyclooxygenase (COX), proteins implicated in prostaglandin production, which contribute to tumorigenesis and metastasis by promoting inflammation and angiogenesis [43,44]. Their PCM model incorporated 11 mammalian COX amino acid sequences and chemical descriptors for over 3200 compounds. They implemented an ensemble machine learning model to predict inhibitor potency using IC50 values derived from ChEMBL with good accuracy and were further able to identify sub-structure chemical moieties implicated in potency and selectivity. Kundu et al. applied Random Forest PCM machine learning to predict binding affinity using data from the PDBbind Database [45] with input features from proteins (including sequence and secondary structure information) and ligands (including physicochemical properties) [46].

Deep learning approaches incorporating protein and ligand chemical features for PCM modeling have also recently emerged. DeepDTA is a CNN-based drug-target affinity prediction model that learns from just sequence information of proteins and compounds (amino acids and SMILES, respectively) [47]. This is notable in that DeepDTA, using only raw sequences as input, performed as well as other established machine

learning methods that implement multiple tools and algorithms. Cui et al. established DeepCSeqSite based on deep convolutional neural networks for predicting protein-ligand binding residues using the BioLip database [48] for model training [49]. PCM deep learning also has utility for enriching docking results through refinement of scoring functions and increasing ligand pose prediction accuracy. For example, Jimenez et al. employed 3D-convolutional neural networks for predicting binding affinity from three-dimensional ligand-protein structures obtained from PDBbind [50].

6. Machine learning on cell phenotypic data

Although associating a target's inhibition with a desired clinical outcome has been successful in multiple cases, our limited knowledge of the complete interactome prohibits any systematic target-phenotype extrapolations. This, however, can be partially overcome by studying the direct phenotypic effects of drugs on cells and complex tissues with dose-response, cell death and apoptosis assays. Moreover, the ability to easily scale such assays into a high-throughput format allows the generation of large-scale datasets and enables the utilization machine and deep learning approaches for the prediction of drug-phenotype associations (Fig. 2).

These high-throughput phenotypic datasets are in most cases publicly accessible through dedicated data repositories and portals that enable the researchers to quickly access, explore and analyze them. One of the largest and widely used repositories is the PubChem bioassay repository [51]. As of 2017, PubChem contained over 1.2 million assays comprising ~2.2 million chemicals, ~65,000 protein/gene targets and over 3000 species comprising ~230 million bioactivity outcome data points. ChEMBL [25] is also another large-scale chemical bioactivity repository with. Publicly available tools such as DPubChem [52] have also been developed to allow for streamlining machine learning applications for investigators.

Developing machine learning models based on the above large-scale datasets has resulted in interesting new drug repurposing opportunities. One such example is the study from Kadurin et al. in which they applied adversarial autoencoders to full dose response data from the NCI-60 cell

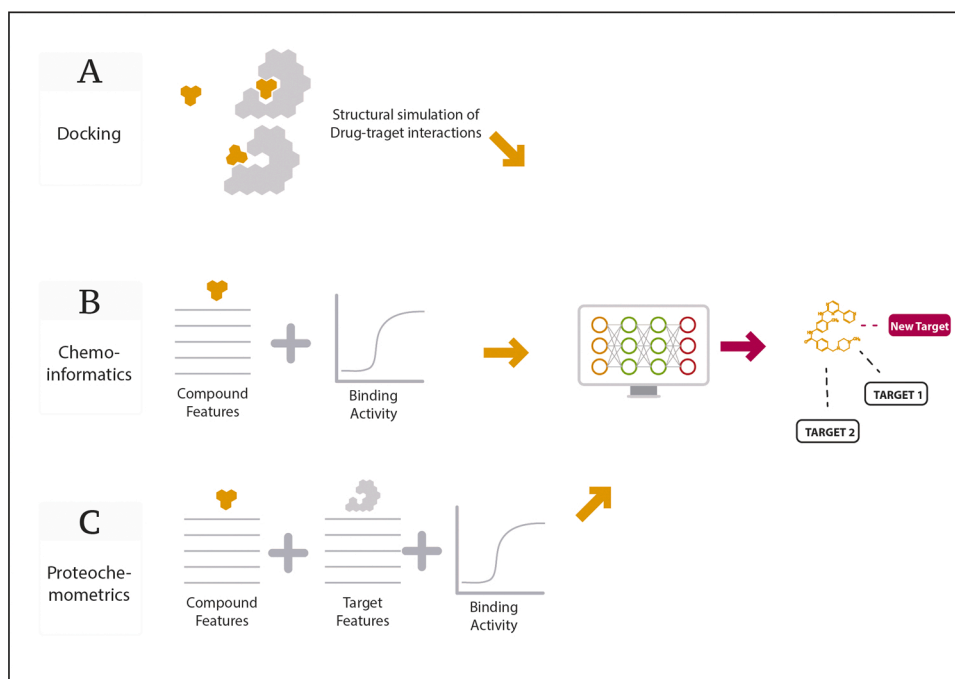


Fig. 2. Schematic of machine learning applied to (A) phenotypic and (B) transcriptomic data. Bioassay data can be obtained from numerous publicly available databases such as PubChem, ChEMBL, TCGA and L1000 and implemented into machine learning systems.

line assay to develop a deep learning model [53]. They then pipelined the PubChem database of over 72 million compounds and were able to validate their model by finding molecules with already proven anti-cancer activity as well as revealed new molecules with potential anti-cancer activity, but had not yet been studied for that indication.

High throughput imaging assays also serve as rich sources of phenotypic data in drug screenings. For example, morphological changes of cells exposed to drugs may be captured via cell staining microscopy and morphology information can be extracted using software such as CellProfiler [54] and transformed into image-based compound fingerprints amenable to machine learning for the prediction of compound activity [55]. In a recent study, Simm et al. created a high-dimensional image-based fingerprint for compound-treated cells in a three-channel screen for glucocorticoid receptor translocation [56]. 524,371 compounds were screened against H4 brain neuroglioma cells and 842-dimensional vectors for each compound were extracted to form fingerprints. These fingerprints, combined with known drug bioactivity data from orthogonal assays (~1200 bioassays across all screened compounds), were used as input for different machine and deep learning protocols.

Machine learning applied to chemical phenotypic screens may also conversely help to identify new disease-relevant targets - a process known as “target deconvolution”. This is possible through exploitation of drug polypharmacology. Using diverse libraries of drugs with known target interaction profiles (*i.e.* inhibitory and binding constant values), one can screen a library to see effects on a particular phenotype (*i.e.* cell growth) and further apply machine learning to find important targets. Gujral et al. utilized this approach to identify cancer cell type-specific kinases relevant to cell migration [57]. They first created an optimized library of 32 kinase inhibitors that cover a significant portion of the kinome and screened those molecules against six different cancer cell types in real time wound closure assays using 96-well plates. EC50 (50 % reduction in cell migration) values at 12-h time points were obtained. They then employed elastic net regularization on the EC50 values and drug-target interaction map to narrow down the list of candidate kinases to less than 30 for each cancer cell type. To further distinguish which kinases are causal to the cell migration phenotype from non-causal or even non-expressed kinases, they subsequently used mRNA expression profiling to further prune the list of kinases and performed gene knockdown experiments for validation. Al-Ali et al. similarly used support vector machines (SVMs) to relate kinase inhibition data to neurite growth - a process critical for neural regeneration in central nervous system injury [58]. Using SVMs, they were not only able to identify kinase targets whose inhibition promoted neurite growth, but also found kinase “anti-targets” whose inhibition blocked growth. They further found growth-promoting synergy among a 2-drug combination that collectively covers the whole span of robust targets. Their SVM model, termed idTRAX, was recently applied to triple negative breast cancer cell lines to identify cell line-specific kinases critical for cell survival (*i.e.* inhibition of FGFR2 specifically kills MFM-223 cells) [59]. Coupling target deconvolution methods for cancer cell-specific phenotypes with the fast growing datasets of drug-target interaction profiles will further enable novel precision drug repurposing opportunities.

7. Transcriptomics-based machine learning

The advent of Next Generation Sequencing (NGS) technologies have enabled the cost-effective, reliable and large-scale generation of transcriptional profiles that provide considerable insights into the complex and aberrant regulatory network in different cancers. The Cancer Genome Atlas (TCGA) has been at the forefront of this NGS data generation effort, providing the largest uniformly processed multi-omic cancer dataset from 33 tumor types [60]. Gene expression profiles obtained from thousands of patient samples deposited in the TCGA, as well as other public databases such as the NCBI Gene Expression Omnibus [61], have enabled the adoption of comprehensive systems biology

approaches that together with machine learning approaches can elucidate new disease target and next generation treatments.

In addition to finding pharmacologically relevant cancer targets, large-scale transcriptomics datasets can be used to delineate drug mechanisms of action and uncover drug repurposing opportunities (Fig. 2). The Connectivity Map (CMap) approach is among the most well-known approaches where drugs that induce discordant gene expression profiles to a disease-specific gene expression profile are likely to be therapeutic candidates for that disease [63]. Over the past years, connectivity mapping has been successful in illuminating unforeseen cancer drug repurposing opportunities. Sirota et al. predicted and validated cimetidine, an off-patent approved anti-ulcer drug with favorable safety profile, to be indicated for lung adenocarcinoma [64]. This effect was replicated recently in xenograft model of A549 lung adenocarcinoma cells [65]. Jahchan et al. also identified tricyclic antidepressants to inhibit small cell lung cancer as well as neuroendocrine tumors [66]. Previous success of CMap in cancer [67] and other disease states [68–71] led to a significant expansion of the reference drug perturbation gene expression dataset to increase the likelihood of uncovering new drug indications. A major effort was established by the NIH LINCS Consortium [72,73] to establish a 1,000-fold scale-up of the original CMap using a novel gene expression profiling method called the L1000 that is of low cost and high throughput [74]. Using the L1000 assay, over 42,000 unannotated perturbagens (including drugs, gene knockdowns and gene overexpressions) were screened against 77 cell lines to create over a million gene expression perturbation signatures. These transcriptional response signatures have helped describe and elucidate the effects of drug treatments on the transcriptional level and have been instrumental in the cancer drug repurposing opportunities, such as emetine, a pro-emetic, for atypical meningiomas [75]. Furthermore, apart from single drug predictions, the use of the L1000 data has also aided in the development of methods for predicting synergistic drug combinations such as the SynergySeq [76] and SynGeNet [77] platforms.

In addition to the typical signature anti-correlated matching, the large amount of data generated from L1000 is ideal for machine learning approaches to recapitulate and find novel drug-target interactions. Xie et al. applied a supervised deep neural network (DNN) to L1000 gene expression data and a drug-target interaction activity dataset from DrugBank [78] and accurately predicted established drug-target pairs using external validation sets [79]. Filzen et al. similarly utilized the Pylearn2 deep learning framework for model building from the L1000 along with a different drug-target annotation dataset obtained from Merck (CHEMGENTIE) [80]. Chiu et al. used DNN combining inputs from TCGA mutation linked to cancer-specific gene expression data and relating them to drug response [81]. Their DNN model, called DeepDR, predicted IC50 values for 265 drugs and found a novel use for the chemotherapeutic drug vinorelbine in TTN-mutated tumors. Zhang et al. integrated TCGA copy number data in addition to mutation and gene expression data to develop their deep learning model, AuDNNsynergy, to predict synergistic drug combinations [82]. Zhou et al. designed a gene expression-weighted cosine method called EMUDRA from Connectivity Map and LINCS data and experimentally validated the repurposing of the antibiotic rifabutin as an inhibitor of triple negative breast cancer cell growth [83]. Combining drug perturbation gene expression with that of cancer cell lines and pathway activation data, Aliper et al. used DNNs and predict the potassium channel-blocking vasodilator drug pinacidil as a central nervous system (CNS) agent [84] with the potential to be repurposed for poor-prognosis CNS cancers such as glioblastoma [86].

8. EHR-based machine learning

In addition to the utilization of cheminformatic, phenotypic and transcriptomic datasets, electronic health records (EHRs) also contain a wealth of information that can be mined for drug discovery, repurposing

and adverse drug event (ADE) predictions. Recent advances in natural language processing (NLP) has enabled investigators to take the largely unstructured data sources (*i.e.* clinic notes, web forum posts) and translate them into a machine-interpretable format, thus enabling the application of machine and deep learning analytics. For example, in an effort to improve upon the underreporting of ADEs through current means such as the FDA Adverse Events Reporting System, Li et al. employed a deep learning model on EHR notes to more accurately detect ADEs than the most current state-of-the-art [87].

An important drug repurposing opportunity comes from off-label use data of approved drugs. Both EHR and social media data have been successfully used to determine off-label uses. In one such study, Jung et al. built a training set of known drug usage using the Medi-Span Drug Indications Database and used it to train a support vector machine (SVM) classifier that was able to discriminate between known drug use indications and off-label use from annotated clinical text, thus producing 10,765 potential novel drug-indication pairs that could be used for EHR-based drug repurposing [88]. In another EHR-based study, Xu et al. developed a machine learning framework that led to the association between metformin, an anti-diabetic medication, and improved cancer survival rates [89]. By integrating two EHRdatabases from the Mayo Clinic and Vanderbilt University Medical Center to their tumor registries, they generated two independent cohorts totaling over 100,000 patients between 1995–2010. While controlling for other covariates such as smoking, metformin was found to be associated with decreased mortality after a cancer diagnosis. The repurposing potential for metformin as an anti-cancer therapeutic has also been confirmed in several preclinical and epidemiological studies for breast cancer [90].

As NLP and deep learning algorithms increase in sophistication with a growing body of clinical data available from EHR and online websites, we imagine that other cancer drug repurposing instances will arise from similar endeavors. There has also been an increased involvement of large private entities such as Google and Amazon that have developed their own deep learning platforms (Google AutoML and Amazon Comprehend Medical) and application programming interface (APIs) [91]. With involvement of these entities and their supercomputing capabilities, identification of new drug repurposing opportunities through NLP/deep learning will undoubtedly be accelerated.

9. Drug repurposing and immunotherapy

A major property of poor prognosis malignancies is their ability to evade detection and killing by the patient's immune system [92]. The tumor's intrinsic biology and its immune microenvironment both play critical roles in tumor immune evasion. The relative proportions and absolute amounts of different immune cell types within the tumor microenvironment determine a tumor's ability for immune evasion. For example, dominant presence of regulatory T cells confers an immunotolerant phenotype. A complex interplay of immune cell types, expressed antigens and secreted factors alters the ability of the immune system to eradicate tumor cells. A detailed discussion of the mechanisms of immune evasion can be found in the following reference [94].

Much attention has been placed on immunomodulation of the tumor microenvironment for therapy. CD4⁺ FoxP3⁺ T regulatory cells (Tregs), which provide a suppressive effect on effector T cells implicated in tumor surveillance, have been shown to represent a large proportion of tumor-infiltrating lymphocytes (TILs) in melanoma and breast cancer immune checkpoint inhibitor (ICI) non-responders [102]. Thus, the attenuation or abolition of tumor-infiltrating Treg function has been a subject of great interest in immuno-oncology therapeutics [103]. However, a major concern of Treg depletion is increased immune-related adverse events such as induction of autoimmune disorders [105]. In lieu of depletion, biologic- and small molecule-driven alteration of Treg function is under intense study [106]. The long-established chemotherapeutic cyclophosphamide was recently found to inhibit Tregs at a low dose thus posing an effective drug repurposing opportunity

immunotherapy [107]. Wang et al. found Treg function in tumor infiltrates to be dependent on the enhancer of zeste homolog 2 (EZH2) protein was also identified [108]. Pharmacological inhibition of EZH2 resulted in Tregs acquiring a pro-inflammatory function that enhanced anti-cancer immunity. Thus, application of machine and deep learning methods to find novel inhibitors of EZH2 may be a promising strategy for repurposing drugs for Treg-driven immunomodulation.

In addition to direct regulation of Tregs, other immune or tumor cell protein targets implicated in immunomodulation are of pharmaceutical interest. Indoleamine 2,3-dioxygenase (IDO) is a target with a multifaceted effect on tumor microenvironment immunosuppression. IDO catalyzes the degradation of tryptophan along the kynurenine pathway [109]. Depletion of tryptophan in the tumor microenvironment consequently activates the GCN2 kinase pathway in neighboring cells, thus biasing naive CD4⁺ T cells to differentiate into Treg cells [110,111]. In addition, tryptophan metabolites (*e.g.* kynurenine) generated from IDO catalysis bind aryl hydrocarbon receptor (AHR) to induce Treg cells and a tolerogenic phenotype of dendritic cells [112]. Apoptotic tumor cells also stimulate IDO expression in neighboring cells and can contextually alter the balance between pro-inflammatory and immunosuppressive states in the microenvironment [113]. In addition to IDO, the adenosine A2A receptor is yet another target of interest for immunomodulation [114]. Elevated adenosine levels have been recognized in the tumor microenvironment resulting in an immunosuppressive effect on NK and T cells [115,116]. Antagonism of the A2A receptor is thus an additional viable strategy for immune checkpoint blockade [117].

Application of virtual screenings and machine learning has helped identify novel inhibitors of IDO and A2AR. Zhang et al. took a library of known IDO inhibitors and developed naive Bayesian and recursive partitioning models that identified three new IDO inhibitors with low-micromolar inhibitory concentrations and activity in cell-based assays [118]. Interestingly, these inhibitors were of the tanishone family with molecular scaffolds derived from the Chinese herb Danshen, which is widely used for cardiovascular and cerebrovascular diseases. Zhou et al. employed a ligand-based 3D-QSAR using pharmacophore models derived from known IDO inhibitors to find novel backbones for next generation IDO inhibitors [119]. SVMs applied to the repertoire of A2AR inhibitors helped identify novel pyrazolo-triazolo-pyrimidine derivatives with A2AR activity [120]. SVMs were also used by Shao et al. to identify novel compounds with the ability to simultaneously inhibit A2AR and dopamine D2 receptor at low-micromolar concentrations [121]. Liu et al. applied recurrent neural networks (RNN) to design a library of novel A2AR inhibitors that exhibited large diversity in their coverage of the chemical space [122]. Tian et al. performed docking virtual screening of over 2.65 million compounds against A2AR crystal structures to find a novel A2AR inhibitor with nanomolar binding affinity [123].

The vascular endothelial growth factor (VEGF)-VEGFR2 axis is also of great pharmacologic interest in cancer therapy. Inhibition of this pro-angiogenic pathway through VEGFR2 kinase inhibition has been efficacious for numerous cancers [124] by reducing immunosuppression by preventing recruitment of immature dendritic cells, myeloid-derived suppressor cells and Tregs [125,126]. Anti-angiogenic agents targeting this axis promote a pro-inflammatory microenvironment [127]. Drug repurposing efforts to inhibit VEGFR2 have been underway. Our group previously employed proteochemometrics to identify the approved anti-helmenthic drug mebendazole as a VEGFR2 inhibitor and anti-angiogenic [128]. Kang et al. developed a naive Bayesian machine learning model using a training set of 2598 inhibitors and 7764 decoys and screened 1841 FDA-approved drugs [129]. They discovered flubendazole, rilpivirine and papaverine to inhibit VEGFR2 with sub- to low-micromolar IC50 values.

IL-6 is another immunomodulatory target shown to suppress T cell-mediated anti-tumor immunity [130]. IL-6 blockade, in combination with other treatment modalities, provides a beneficial anti-tumor response [131]. As IL-6 binds GP130, Chen et al. employed a novel

combinatorial approach to target GP130 and found the FDA-approved drug bazedoxifene, indicated for osteoporosis, to block that interaction and inhibit the increased proliferation of triple-negative breast cancer cells induced by IL-6 [132]. Han et al. also found the molecule CGP-60474, a CDK inhibitor, to reduce IL-6 levels in murine macrophage cell lines [133]. They applied Connectivity Mapping on sepsis-related microarray gene expression studies using the L1000 perturbation set and found CGP-60474 to reverse the sepsis gene signature. Subsequent analysis of the anti-inflammatory effect of CGP-60474 revealed it to inhibit IL-6 levels.

Macrophage polarization has also been implicated in tumor biology. M2/repair-type macrophages promote tumor growth whereas M1/kill-type proinflammatory macrophages halt growth [134]. High M1/M2 macrophage ratio decreases cancer susceptibility, and modulation of M2 to M1 phenotype (or alteration of the M1/M2 ratio) seems to be a viable immunotherapy modality. Some targetable proteins involved in the modulation of macrophage polarization, such as the TRPV4 ion channel [135] and cannabinoid receptor 2 [136], have already been elucidated. Structure based approaches noted above could be leveraged to identify drug repurposing opportunities against those targets. In addition, some drugs have also been shown to alter macrophage polarization such as praziquantel [137] and capsaicin [138]. Machine and deep learning methods applied to these drugs in conjunction with their quantifiable phenotypic effect on macrophages could further help uncover unrelated drugs with similar capabilities.

10. Considerations in machine learning- and AI-Driven methods for drug repurposing

The accuracy of any proposed machine learning- or AI-driven model is contingent upon the quality, type and amount of data in the training set as well as the validation methods of those models. With respect to the former, the training set dictates the applicability domain of the model the query molecule space from which the model can make accurate predictions [139]. For example, if a model is created from a small set of inhibitors for a target containing the same chemical backbone, then the model would be successful in predicting new active compounds with similar backbones but not actives with diverse structures. This narrowness in a model's applicability domain is also known as overfitting.

Deep learning approaches are particularly prone to overfitting, especially as a network's complexity increases. Numerous methods help to reduce overfitting including: (1) making the network smaller (removing layers) (2), neuron dropout regularization (ignoring neuron units during the training phase at random), and (3) early stopping (termination of training at the junction where validation performance starts to deteriorate) [140,141]. Rifaioğlu et al. implemented neuron dropout regularization in their model training of DEEPScreen, validated using Monte Carlo Cross-validation, to predict drug-target interactions from two-dimensional structural images of drugs [33]. Chiu et al. utilized the early stopping method in the development of their highly complex deep neural network that integrates drug, TCGA tumor mutational data and CCLE tumor gene expression data that predicts tumor response to drugs [81].

Type and number of input descriptors also affect model performance. Typically, protein and ligand structural, chemical and physical features are represented as “fingerprints” - vectors of varying sizes containing 0's or 1's that numerically represent the absence or presence of a feature, respectively. Fingerprints can be used to represent both two-dimensional (e.g. atom pairs, atom types, linear atom connectivity) and three-dimensional features (e.g. neighboring covalently bound and unbound atoms and atom functional group categories in 3D space). Thousands of fingerprints exist and have been extensively used in cheminformatics, and their relatedness and performance has been evaluated [142]. In addition to fingerprints, integers and floating numbers can be used to represent properties such as molecular weight or solvent-accessible surface area. Other potential descriptor inputs include

biological data such as mutations, gene expression and others. Thus, a variety of descriptor inputs exists for model generation.

The choice of input descriptor depends on the prediction problem. For example, pharmacophore fingerprints may work better for discovering new chemotypes. Pharmacophores are vectorized 3D representations of chemical functional groups (e.g. hydrophobic centers, hydrogen bond donors and acceptors, etc.) and the distances between them necessary to establish specific ligand-target interactions. As pharmacophores are more abstract representations of atom types, they can be used to screen for structurally diverse ligands with the ability to recapitulate similar interactions to the same target binding site. Use of pharmacophores in machine learning-based predictions is exemplified by Taha et al. [30]. They generated pharmacophore models and other 2D descriptor fingerprints from a known set of PI3Kgamma inhibitors and created a QSAR model of the optimal combination of pharmacophores/2D descriptors using IC50 values as the biological endpoint. The resulting QSAR model was then used to screen the National Cancer Institute list of >200,000 compounds and identified 9 compounds with diverse chemotypes and sub-micromolar IC50 values against PI3Kgamma. Alam et al. also generated pharmacophore models of maslinic acid analogs and subsequently subjected them to QSAR modeling using their IC50 inhibitor concentrations against the MCF-7 breast cancer cell line [28]. The ZINC database was then screened using the developed QSAR model resulting in the identification of compound P-902 with potential anticancer activity against MCF-7 breast cancer cells. Zhou et al. similarly employed QSAR modeling of IDO1 inhibitor pharmacophores to discover novel backbones for next generation IDO inhibitors [119].

In addition to pharmacophore fingerprints, circular fingerprints (e.g. ECFP) also offer a good input choice that can exhibit good performance. ECFPs are fingerprints that are iteratively generated by first assigning a feature to a selected atom in a molecule and then radially expanding to neighboring atoms [143]. These features may represent atom types, bond connectivity, atomic weight, atomic charge, number of attached hydrogens, pharmacophoric property, among many others. The number of iterations is up to the discretion of the user. Structural protein-ligand interaction fingerprints are also widely used in proteochemometrics as they numerically represent three-dimensional structural information derived from biologically relevant protein-ligand interactions [144, 145]. Interaction fingerprints could be considered as protein residues within a certain distance of each ligand atom or the center of the entire ligand obtained from crystallography or structural molecular modeling. Like ECFPs, interaction fingerprints are successful at predicting new ligands for a target. For example, van Westen et al. implemented both ligand-based circular and interaction fingerprints to predict novel adenosine receptor ligands [42]. They found six novel high-affinity ligands, two of which resemble structures of known adenosine receptor binders while others exhibited diverse chemistry. Interestingly, the combination of circular and interaction fingerprints was able to uncover new chemotypes while also prioritizing compounds with similar chemotypes. Another descriptor used by Rifaioğlu et al. is the actual two-dimensional structural image of drugs as input for their CNN [33]. Focused on gene expression-driven inputs, Aliper et al. utilized transcriptional drug perturbation responses across different cell types as inputs for their deep neural network [84]. They found that pathway-level information performed significantly better than gene level data due to reduction of dimensionality while keeping biological relevance. Choice and applicability of data types and fingerprints used in machine learning- and AI-driven methods are highly nuanced to the problem posed. There usually is an iterative process in model development where different descriptors/combinations are tried such that performance is maximized, which changes across desired biological endpoints and targets.

Dataset size is also a crucial aspect of in machine learning. In general, more data incorporated into the training set generally results in better model performance (less overfitting), especially for deep neural

networks. For example, Allen et al. found that for predicting drug-kinase activity using deep learning, model performance was significantly less when the number of active compounds for a kinase was less than 500 and further degraded when there was less than 50 [31]. This is with respect to single-task learning (*i.e.* predicting drug activity to a single kinase). Interestingly, they found that their drug-kinase multi-task deep neural network, where they predicted drug-kinase activity across 342 kinases simultaneously using a well-curated aggregate dataset, continued to perform well even with low number of actives. Their model was further validated by 5-fold cross validation.

Data quality is also of critical importance. Generally, data inputted into models are typically generated from high-throughput screens. For example, PubChem [23] contains information from biological assays for drug-target binding events (*e.g.* binding affinity) or drug-induced changes in gene expression, phosphorylation as well as cell viability. Such assays are performed by various groups under different conditions, and data is being generated at an exponential pace. Thus, in order to more optimally integrate this data into machine learning- and AI-assisted drug discovery efforts, the data needs conform to community standards of findability, accessibility, interoperability and reusability (FAIR standards) [146,147]. Metadata (*e.g.* data regarding experimental conditions, species studied, tissue obtained, *etc.*) is of critical importance as it describes the nature of the conditions under which the data was collected and provides further context from which the model can learn. For example, biological phenotypes due to drugs such as estrogen receptor transcriptional response may be different across cell lines due to differences in expressed co-activators, co-repressors, *etc.* Such information can be captured by well-annotated metadata and help fine-tune models for greater accuracy. Thus, appropriate data management with standards becomes a priority for drug discovery data generation [148]. Our group has formally defined the details of bioassays [149–151] and drug targets [152] in an effort to lay the foundations for future data generation and analytics. By upholding community-accepted data standards, poor quality data can easily be filtered out prior to model generation, and data generated in the future will conform to the same standards and thus be easily integrated in various machine-learning and AI-driven drug discovery efforts while minimizing the need for extensive manual curation. As the amount of data continues to exponentially increase along with computing capabilities, we anticipate that future modeling efforts will be exceedingly more accurate and uncover new drug repurposing opportunities for cancer therapy.

11. Conclusion

With massive data collection efforts such as the TCGA, our understanding of tumor pathogenesis and survival has evolved significantly. The implementation of newer “omics” technologies such as proteomics and metabolomics lend new insights into cancer cell and immunological processes contributing to disease progression [153–155]. The movement toward single cell RNA sequencing has already provided further perspectives into tumor heterogeneity at a highly granular level not previously appreciated from bulk biopsy transcriptomics [156]. As these technologies, mature, novel pharmacologically actionable targets will further be uncovered.

At this time, a major limitation in the success of machine and deep learning applications for drug repurposing is the quality and quantity of available data. As the costs of highly automated high-throughput screening platforms continue to decrease, the amount of data produced from drug assay screening is increasing exponentially. Rigorous analytical frameworks for annotating drug discovery data (*i.e.* ontologies) are also necessary for true standardization and interoperability of data produced and integration into machine learning programs [151, 152,157,158]. Further, model applicability will be enhanced with more data, better data integration, the true drug-target interaction space is expanding significantly, and the full extent of drug polypharmacology is starting to emerge. Such a large dataset increases its drug domain

coverage that could bring equal numbers of classes to the training dataset potentially reducing the class imbalance problem in the machine learning models. The physiological effects of drugs on different cell phenotypes are better understood via increasingly thorough multi-omics profiling and sophisticated systems biology models. Coupled with the advancement of computational drug discovery strategies such as machine learning and artificial intelligence (deep learning), this information can be leveraged to identify new drug-target and drug-cell interactions.

As the development of new molecular entities is risky and suffers from a high cost burden, drug repurposing of approved and investigational drugs has been touted as a strategy with lower barrier to success. The clinical potential of drug repurposing in oncology has yet to be realized at this time. With the ever-increasing amount of biological and pharmacological data, computational methods will certainly provide new perspectives into therapeutic mechanisms and enable drug repurposing within the framework of precision oncology.

Transparency document

The [Transparency document](#) associated with this article can be found in the online version.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Acknowledgements

This work was supported by NIH grants U54HL127624 (BD2K LINC Data Coordination and Integration Center, DCIC), U24TR002278 (Illuminating the Druggable Genome Resource Dissemination and Outreach Center, IDG-RDOC), and U01LM012630 (BD2K, Enhancing the efficiency and effectiveness of digital curation for biomedical ‘big data’), and the State of Florida Biomedical Research Program, Bankhead Coley grant number 9BC13. The author Sivanesan Dakshanamurthy wishes to acknowledge the support in part by DOD grant CA140882, CCSG grant NIH-P30 CA051008, GUMC Computational Chemistry Shared Resources (CCSR), and GUMC Lombardi Comprehensive Cancer Center.

References

- [1] S. Ekins, A.C. Puhl, K.M. Zorn, T.R. Lane, D.P. Russo, J.J. Klein, et al., Exploiting machine learning for end-to-end drug discovery and development, *Nat. Mater.* 18 (5) (2019) 435–441.
- [2] N. Nosengo, Can you teach old drugs new tricks? *Nature* 534 (7607) (2016) 314–316.
- [3] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.* 47 (7) (2004) 1739–1749.
- [4] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31 (2) (2010) 455–461.
- [5] Y. Wang, H.Q. Lin, P. Wang, J.S. Hu, T.M. Ip, L.M. Yang, et al., Discovery of a Novel HIV-1 Integrase/p75 Interacting Inhibitor by Docking Screening, Biochemical Assay, and in Vitro Studies, *J. Chem. Inf. Model.* 57 (9) (2017) 2336–2343.
- [6] S.B. Mirza, R.E. Salmas, M.Q. Fatmi, S. Durdagi, Virtual screening of eighteen million compounds against dengue virus: combined molecular docking and molecular dynamics simulations study, *J. Mol. Graph. Model* 66 (2016) 99–107.
- [7] V. Kumar, S. Krishna, M.I. Siddiqi, Virtual screening strategies: recent advances in the identification and design of anti-cancer agents, *Methods*. 71 (2015) 64–70.
- [8] B.B. Hafeez, A. Ganju, M. Sikander, V.K. Kashyap, Z.B. Hafeez, N. Chauhan, et al., Ormeloxifene suppresses prostate tumor growth and metastatic phenotypes via inhibition of oncogenic beta-catenin signaling and EMT progression, *Mol. Cancer Ther.* 16 (10) (2017) 2267–2280.
- [9] Y.C. Chen, Beware of docking!, *Trends Pharmacol. Sci.* 36 (2) (2015) 78–95.
- [10] I. Wallach, M. Dzamba, Heifets A. AtomNet, A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery, 2015 October 01 Available from: arXiv e-prints [Internet], 2015 <https://ui.adsabs.harvard.edu/abs/2015arXiv151002855W>.

- [13] J. Desaphy, G. Bret, D. Rognan, Kellenberger E. sc-PDB: a 3D-database of ligandable binding sites–10 years on, *Nucleic Acids Res.* 43 (Database issue) (2015) D399–404.
- [14] R. Spitzer, A.N. Jain, Surflex-Dock: docking benchmarks and real-world application, *J. Comput. Aided Mol. Des.* 26 (6) (2012) 687–699.
- [15] W.J. Allen, T.E. Balias, S. Mukherjee, S.R. Brozell, D.T. Moustakas, P.T. Lang, et al., DOCK 6: impact of new features and current docking performance, *J. Comput. Chem.* 36 (15) (2015) 1132–1156.
- [16] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D.R. Koes, Protein-ligand scoring with convolutional neural networks, *J. Chem. Inf. Model.* 57 (4) (2017) 942–957.
- [17] Y. Li, L. Han, Z. Liu, R. Wang, Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results, *J. Chem. Inf. Model.* 54 (6) (2014) 1717–1736.
- [18] D.D. Nguyen, Z. Cang, K. Wu, M. Wang, Y. Cao, G.W. Wei, Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges, *J. Comput. Aided Mol. Des.* 33 (1) (2019) 71–82.
- [19] A.G. Maldonado, J.P. Doucet, M. Petitjean, B.T. Fan, Molecular similarity and diversity in chemoinformatics: from theory to applications, *Mol. Divers.* 10 (1) (2006) 39–79.
- [20] M.J. Keiser, V. Setola, J.J. Irwin, C. Laggner, A.I. Abbas, S.J. Hufeisen, et al., Predicting new molecular targets for known drugs, *Nature.* 462 (7270) (2009) 175–181.
- [21] Y. Hu, D. Stumpfe, J. Bajorath, Advancing the activity cliff concept, *Fl000Res.* 2 (2013) 199.
- [22] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology, *Nucleic Acids Res.* 44 (D1) (2016) D1045–53.
- [23] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, et al., PubChem 2019 update: improved access to chemical data, *Nucleic Acids Res.* 47 (D1) (2019) D1102–D1109.
- [24] J.M. Barbarino, M. Whirl-Carrillo, R.B. Altman, T.E. Klein, PharmGKB: a worldwide resource for pharmacogenomic information, *Wiley Interdiscip. Rev. Syst. Biol. Med.* 10 (4) (2018) e1417.
- [25] A. Gaulton, A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, et al., The ChEMBL database in 2017, *Nucleic Acids Res.* 45 (D1) (2017) D945–D954.
- [26] A.L. Deshmukh, S. Chandra, D.K. Singh, M.I. Siddiqi, D. Banerjee, Identification of human flap endonuclease 1 (FEN1) inhibitors using a machine learning based consensus virtual screening, *Mol. Biosyst.* 13 (8) (2017) 1630–1639.
- [27] Z.Y. Algalal, M.H. Lee, A.M. Al-Fakih, M. Aziz, High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO, *J. Chemom.* 29 (10) (2015) 547–556.
- [28] S. Alam, F. Khan, 3D-QSAR studies on Maslinic acid analogs for Anticancer activity against Breast Cancer cell line MCF-7, *Sci. Rep.* 7 (1) (2017) 6019.
- [29] T. Sterling, J.J. Irwin, ZINC 15–Ligand discovery for everyone, *J. Chem. Inf. Model.* 55 (11) (2015) 2324–2337.
- [30] M.O. Taha, M.A. Al-Sha'er, M.A. Khanfar, A.H. Al-Nadaf, Discovery of nanomolar phosphoinositide 3-kinase gamma (PI3Kgamma) inhibitors using ligand-based modeling and virtual screening followed by in vitro analysis, *Eur. J. Med. Chem.* 84 (2014) 454–465.
- [31] B.K. Allen, N.G. Ayad, S.C. Schürer, Kinome-wide activity classification of small molecules by deep learning, *bioRxiv.* 512459 (2019).
- [32] S.C. Schurer, S.M. Muskul, Kinome-wide activity modeling from diverse public high-quality data sets, *J. Chem. Inf. Model.* 53 (1) (2013) 27–38.
- [33] A.S. Rifaioglu, V. Atalay, M.J. Martin, R. Cetin-Atalay, T. Doğan, DEEPScreen: high performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations, *bioRxiv* (2018), 491365.
- [34] W.C. Cho, OncomiRs: the discovery and progress of microRNAs in cancers, *Mol. Cancer* 6 (2007) 60.
- [35] S. Jamal, V. Periwal, Open Source Drug Discovery C, Scaria V. Computational analysis and predictive modeling of small molecule modulators of microRNA, *J. Cheminform.* 4 (1) (2012) 16.
- [36] X. Liu, S. Wang, F. Meng, J. Wang, Y. Zhang, E. Dai, et al., SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression, *Bioinformatics* 29 (3) (2013) 409–411.
- [37] C.C. Wang, X. Chen, J. Qu, Y.Z. Sun, J.Q. Li, RFSMMA: a new computational model to identify and prioritize potential small Molecule-MiRNA associations, *J. Chem. Inf. Model.* 59 (4) (2019) 1668–1679.
- [38] J. Qu, X. Chen, Y.Z. Sun, Y. Zhao, S.B. Cai, Z. Ming, et al., In silico prediction of small Molecule-miRNA associations based on the HeteSim algorithm, *Mol. Ther. Nucleic Acids* 14 (2019) 274–286.
- [39] X. Chen, N.N. Guan, Y.Z. Sun, J.Q. Li, J. Qu, MicroRNA-small molecule association identification: from experimental results to computational models, *Brief Bioinform.* (2018).
- [40] S. Gessi, S. Merighi, V. Sacchetto, C. Simioni, P.A. Borea, Adenosine receptors and cancer, *Biochim. Biophys. Acta* 1808 (5) (2011) 1400–1412.
- [41] B. Allard, P.A. Beavis, P.K. Darcy, J. Stagg, Immunosuppressive activities of adenosine in cancer, *Curr. Opin. Pharmacol.* 29 (2016) 7–16.
- [42] G.J. van Westen, O.O. van den Hoven, R. van der Pijl, T. Mulder-Krieger, H. de Vries, J.K. Wegner, et al., Identifying novel adenosine receptor ligands by simultaneous proteochemometric modeling of rat and human bioactivity data, *J. Med. Chem.* 55 (16) (2012) 7010–7020.
- [43] I. Cortes-Ciriano, D.S. Murrell, G.J. van Westen, A. Bender, T.E. Malliavin, Prediction of the potency of mammalian cyclooxygenase inhibitors with ensemble proteochemometric modeling, *J. Cheminform.* 7 (2015) 1.
- [44] C. Schneider, A. Pozzi, Cyclooxygenases and lipoxygenases in cancer, *Cancer Metastasis Rev.* 30 (3–4) (2011) 277–294.
- [45] R. Wang, X. Fang, Y. Lu, S. Wang, The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures, *J. Med. Chem.* 47 (12) (2004) 2977–2980.
- [46] I. Kundu, G. Paul, R. Banerjee, A machine learning approach towards the prediction of protein-ligand binding affinity based on fundamental molecular properties, *RSC Adv.* 8 (22) (2018) 12127–12137.
- [47] H. Ozturk, A. Ozgur, Ozkirimli E. DeepDTA, Deep drug-target binding affinity prediction, *Bioinformatics* 34 (17) (2018) i821–i829.
- [48] J. Yang, A. Roy, Zhang Y. BioLiP, A semi-manually curated database for biologically relevant ligand-protein interactions, *Nucleic Acids Res.* 41 (Database issue) (2013) D1096–103.
- [49] Y. Cui, Q. Dong, D. Hong, X. Wang, Predicting protein-ligand binding residues with deep convolutional neural networks, *BMC Bioinformatics* 20 (1) (2019) 93.
- [50] J. Jimenez, M. Skalic, G. Martinez-Rosell, G. De Fabritiis, KDEEP: protein-ligand absolute binding affinity prediction via 3D-Convolutional neural networks, *J. Chem. Inf. Model.* 58 (2) (2018) 287–296.
- [51] Y. Wang, S.H. Bryant, T. Cheng, J. Wang, A. Gindulyte, B.A. Shoemaker, et al., PubChem BioAssay: 2017 update, *Nucleic Acids Res.* 45 (D1) (2017) D955–D963.
- [52] O. Soufan, W. Ba-Alawi, A. Magana-Mora, M. Essack, V.B. Bajic, DPubChem: a web tool for QSAR modeling and high-throughput virtual screening, *Sci. Rep.* 8 (1) (2018) 9110.
- [53] A. Kadurin, A. Aliper, A. Kazennov, P. Mamoshina, Q. Vanhaelen, K. Khrabrov, et al., The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology, *Oncotarget.* 8 (7) (2017) 10883–10890.
- [54] A.E. Carpenter, T.R. Jones, M.R. Lamprecht, C. Clarke, I.H. Kang, O. Friman, et al., CellProfiler: image analysis software for identifying and quantifying cell phenotypes, *Genome Biol.* 7 (10) (2006) R100.
- [55] C. Scheeder, F. Heigwer, M. Boutros, Machine learning and image-based profiling in drug discovery, *Curr. Opin. Syst. Biol.* 10 (2018) 43–52.
- [56] J. Simm, G. Klambauer, A. Arany, M. Steijaert, J.K. Wegner, E. Gustin, et al., Repurposing high-throughput image assays enables biological activity prediction for drug discovery, *Cell Chem. Biol.* 25 (5) (2018) 611–618, e3.
- [57] T.S. Gujral, L. Peshkin, M.W. Kirschner, Exploiting polypharmacology for drug target deconvolution, *Proc. Natl. Acad. Sci. U. S. A.* 111 (13) (2014) 5048–5053.
- [58] H. Al-Ali, D.H. Lee, M.C. Danzi, H. Nassif, P. Gautam, K. Wennerberg, et al., Rational polypharmacology: systematically identifying and engaging multiple drug targets to promote axon growth, *ACS Chem. Biol.* 10 (8) (2015) 1939–1951.
- [59] P. Gautam, A. Jaiswal, T. Aittokallio, H. Al-Ali, K. Wennerberg, Phenotypic screening combined with machine learning for efficient identification of breast cancer-selective therapeutic targets, *Cell Chem. Biol.* (2019).
- [60] Cancer Genome Atlas Research N, J.N. Weinstein, E.A. Collisson, G.B. Mills, K. R. Shaw, B.A. Ozenberger, et al., The Cancer genome atlas pan-cancer analysis project, *Nat. Genet.* 45 (10) (2013) 1113–1120.
- [61] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, et al., NCBI GEO: archive for functional genomics data sets–update, *Nucleic Acids Res.* 41 (Database issue) (2013) D991–5.
- [62] J. Lamb, E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, et al., The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease, *Science* 313 (5795) (2006) 1929–1935.
- [63] M. Sirota, J.T. Dudley, J. Kim, A.P. Chiang, A.A. Morgan, A. Sweet-Cordero, et al., Discovery and preclinical validation of drug indications using compendia of public gene expression data, *Sci. Transl. Med.* 3 (96) (2011), 96ra77–96ra77.
- [64] I. Kandela, F. Aird, Reproducibility Project: cancer B. Replication Study: discovery and preclinical validation of drug indications using compendia of public gene expression data, *eLife.* 6 (2017), e17044.
- [65] N.S. Jahchan, J.T. Dudley, P.K. Mazur, N. Flores, D. Yang, A. Palmerton, et al., A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors, *Cancer Discov.* 3 (12) (2013) 1364–1377.
- [66] H.W. Cheng, Y.H. Liang, Y.L. Kuo, C.P. Chuu, C.Y. Lin, M.H. Lee, et al., Identification of thioridazine, an antipsychotic drug, as an antiglioblastoma and anticancer stem cell agent using public gene expression data, *Cell Death Dis.* 6 (5) (2015) e1753-e.
- [67] A.M. Brum, J. van de Peppel, C.S. van der Leije, M. Schreuders-Koedam, M. Eijken, B.C.J. van der Eerden, et al., Connectivity Map-based discovery of paribendazole reveals targetable human osteogenic pathway, *Proc. Natl. Acad. Sci. U.S.A.* 112 (41) (2015) 12711–12716.
- [68] Y. Ishimatsu-Tsuji, T. Soma, J. Kishimoto, Identification of novel hair-growth inducers by means of connectivity mapping, *Faseb J.* 24 (5) (2010) 1489–1496.
- [69] G. Vanderstocken, A. Dvorkin-Gheva, P. Shen, C.A. Brandsma, M. Obeidat, Y. Bosse, et al., Identification of drug candidates to suppress cigarette smoke-induced inflammation via connectivity map analyses, *Am. J. Respir. Cell Mol. Biol.* 58 (6) (2018) 727–735.
- [70] A.M. Brum, J. van de Peppel, L. Nguyen, A. Aliev, M. Schreuders-Koedam, T. Gajdian, et al., Using the Connectivity Map to discover compounds influencing human osteoblast differentiation, *J. Cell. Physiol.* 233 (6) (2018) 4895–4906.
- [71] A. Kolet, R. Terryn, V. Stathias, C. Chung, D.J. Cooper, J.P. Turner, et al., Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data, *Nucleic Acids Res.* 46 (D1) (2018) D558–D566.
- [72] A.B. Keenan, S.L. Jenkins, K.M. Jagodnik, S. Koplev, E. He, D. Torre, et al., The library of integrated network-based cellular signatures NIH program: system-level

- cataloging of human cells response to perturbations, *Cell Syst.* 6 (1) (2018) 13–24.
- [74] A. Subramanian, R. Narayan, S.M. Corsello, D.D. Peck, T.E. Natoli, X. Lu, et al., A next generation connectivity map: L1000 platform and the first 1,000,000 profiles, *Cell* 171 (6) (2017) 1437–1452, e17.
- [75] Z. Zador, A.T. King, N. Geifman, New drug candidates for treatment of atypical meningiomas: an integrated approach using gene expression signatures for drug repurposing, *PLoS One* 13 (3) (2018) e0194701.
- [76] V. Stathias, A.M. Jermakowicz, M.E. Maloof, M. Forlin, W. Walters, R.K. Suter, et al., Drug and disease signature integration identifies synergistic combinations in glioblastoma, *Nat. Commun.* 9 (1) (2018) 5315.
- [77] J.M. Harrold, M. Ramanathan, D.E. Mager, Network-based approaches in drug discovery and early development, *Clin. Pharmacol. Ther.* 94 (6) (2013) 651–658.
- [78] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.* 46 (D1) (2018) D1074–D1082.
- [79] L. Xie, S. He, X. Song, X. Bo, Z. Zhang, Deep learning-based transcriptome data classification for drug-target interaction prediction, *BMC Genomics* 19 (Suppl 7) (2018) 667.
- [80] T.M. Filzen, P.S. Kutchukian, J.D. Hermes, J. Li, M. Tudor, Representing high throughput expression profiles via perturbation barcodes reveals compound targets, *PLoS Comput. Biol.* 13 (2) (2017) e1005335.
- [81] Y.C. Chiu, H.H. Chen, T. Zhang, S. Zhang, A. Gorthi, L.J. Wang, et al., Predicting drug response of tumors from integrated genomic profiles by deep neural networks, *BMC Med. Genomics* 12 (Suppl 1) (2019) 18.
- [82] T. Zhang, L. Zhang, P.R.O. Payne, F. Li, Synergistic drug combination prediction by integrating multi-omics data in deep learning models, *arXiv e-prints [Internet]* 2018 (November 01) (2018). Available from: <https://ui.adsabs.harvard.edu/abs/2018arXiv181107054Z>.
- [83] X. Zhou, M. Wang, I. Katsyov, H. Irie, B. Zhang, EMUDRA: Ensemble of Multiple Drug Repositioning Approaches to improve prediction accuracy, *Bioinformatics* 34 (18) (2018) 3151–3159.
- [84] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, A. Zhavoronkov, Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data, *Mol. Pharm.* 13 (7) (2016) 2524–2530.
- [86] R.J. Molenaar, Ion channels in glioblastoma, *ISRN Neurol.* 2011 (2011), 590249.
- [87] F. Li, W. Liu, H. Yu, Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning, *JMIR Med. Inform.* 6 (4) (2018) e12159.
- [88] K. Jung, P. Lependu, N. Shah, Automated detection of systematic off-label drug use in free text of electronic medical records, *AMIA Summits Transl. Sci. Proc.* 2013 (2013) 94–98.
- [89] H. Xu, M.C. Aldrich, Q. Chen, H. Liu, N.B. Peterson, Q. Dai, et al., Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality, *J. Am. Med. Inform. Assoc.* 22 (1) (2015) 179–191.
- [90] G.H. Tang, M. Satkunam, G.R. Pond, G.R. Steinberg, G. Blandino, H. J. Schunemann, et al., Association of metformin with breast Cancer incidence and mortality in patients with type II diabetes: a GRADE-Assessed systematic review and meta-analysis, *Cancer Epidemiol. Biomarkers Prev.* 27 (6) (2018) 627–635.
- [91] H.M. Trivedi, M. Panahiazar, A. Liang, D. Lituev, P. Chang, J.H. Sohn, et al., Large scale semi-automated labeling of routine free-text clinical records for deep learning, *J. Digit. Imaging* 32 (1) (2019) 30–37.
- [92] D.S. Chen, I. Mellman, Elements of cancer immunity and the cancer-immune set point, *Nature* 541 (7637) (2017) 321–330.
- [94] D.S. Vinay, E.P. Ryan, G. Pawelec, W.H. Talib, J. Stagg, E. Elkord, et al., Immune evasion in cancer: mechanistic basis and therapeutic strategies, *Semin. Cancer Biol.* 35 (Suppl) (2015) S185–S198.
- [102] N.A. Taylor, S.C. Vick, M.D. Iglesias, W.J. Brickey, B.R. Midkiff, K.P. McKinnon, et al., Treg depletion potentiates checkpoint inhibition in claudin-low breast cancer, *J. Clin. Invest.* 127 (9) (2017) 3472–3483.
- [103] K. Shitara, H. Nishikawa, Regulatory T cells: a potential target in cancer immunotherapy, *Ann. N. Y. Acad. Sci.* 1417 (1) (2018) 104–115.
- [105] J. Liu, S.J. Blake, H. Harjunpaa, K.A. Fairfax, M.C. Yong, S. Allen, et al., Assessing immune-related adverse events of efficacious combination immunotherapies in preclinical models of Cancer, *Cancer Res.* 76 (18) (2016) 5288–5301.
- [106] S. Han, A. Tokar, Z.Q. Liu, P.S. Ohashi, Turning the tide against regulatory T cells, *Front Oncol.* 9 (2019) 279.
- [107] R. Abu Eid, G.S. Razavi, M. Mkrtchyan, J. Janik, S.N. Khleif, Old-school chemotherapy in immunotherapeutic combination in Cancer, A low-cost drug repurposed, *Cancer Immunol. Res.* 4 (5) (2016) 377–382.
- [108] D. Wang, J. Quiros, K. Mahuron, C.C. Pai, V. Ranzani, A. Young, et al., Targeting EZH2 reprograms intratumoral regulatory T cells to enhance Cancer immunity, *Cell Rep.* 23 (11) (2018) 3262–3274.
- [109] D.H. Munn, A.L. Mellor, Indoleamine 2,3 dioxygenase and metabolic control of immune responses, *Trends Immunol.* 34 (3) (2013) 137–143.
- [110] D.H. Munn, M.D. Sharma, B. Baban, H.P. Harding, Y. Zhang, D. Ron, et al., GCN2 kinase in T cells mediates proliferative arrest and anergy induction in response to indoleamine 2,3-dioxygenase, *Immunity* 22 (5) (2005) 633–642.
- [111] F. Fallarino, U. Grohmann, S. You, B.C. McGrath, D.R. Cavener, C. Vacca, et al., The combined effects of tryptophan starvation and tryptophan catabolites down-regulate T cell receptor zeta-chain and induce a regulatory phenotype in naive T cells, *J. Immunol.* 176 (11) (2006) 6752–6761.
- [112] J.D. Mezrich, J.H. Fechner, X. Zhang, B.P. Johnson, W.J. Burlingham, C. A. Bradfield, An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells, *J. Immunol.* (Baltimore, Md: 1950) 185 (6) (2010) 3190–3198.
- [113] T.S. Johnson, T. McGaha, D.H. Munn, Chemo-immunotherapy: role of indoleamine 2,3-Dioxygenase in defining immunogenic versus tolerogenic cell death in the tumor microenvironment, *Adv. Exp. Med. Biol.* 1036 (2017) 91–104.
- [114] D. Dhanak, J.P. Edwards, A. Nguyen, P.J. Tummino, Small-molecule targets in immuno-oncology, *Cell Chem. Biol.* 24 (9) (2017) 1148–1160.
- [115] A. Ohta, E. Gorelik, S.J. Prasad, F. Ronchese, D. Lukashev, M.K. Wong, et al., A2A adenosine receptor protects tumors from antitumor T cells, *Proc. Natl. Acad. Sci. U. S. A.* 103 (35) (2006) 13132–13137.
- [116] R. Sorrentino, A. Pinto, S. Morello, The adenosinergic system in cancer: key therapeutic target, *Oncoimmunology* 2 (1) (2013), e22448.
- [117] R.D. Leone, Y.C. Lo, J.D. Powell, A2AR antagonists: next generation checkpoint blockade for cancer immunotherapy, *Comput. Struct. Biotechnol. J.* 13 (2015) 265–272.
- [118] H. Zhang, W. Liu, Z. Liu, Y. Ju, M. Xu, Y. Zhang, et al., Discovery of indoleamine 2,3-dioxygenase inhibitors using machine learning based virtual screening, *Medchemcomm* 9 (6) (2018) 937–945.
- [119] Y. Zhou, J. Peng, P. Li, H. Du, Y. Li, Y. Li, et al., Discovery of novel indoleamine 2,3-dioxygenase 1 (IDO1) inhibitors by virtual screening, *Comput. Biol. Chem.* 78 (2019) 306–316.
- [120] L. Michielan, C. Bolcato, S. Federico, B. Cacciari, M. Bacilieri, K.-N. Klotz, et al., Combining selectivity and affinity predictions using an integrated Support Vector Machine (SVM) approach: An alternative tool to discriminate between the human adenosine A2A and A3 receptor pyrazolo-triazolo-pyrimidine antagonists binding sites, *Bioorg. Med. Chem.* 17 (14) (2009) 5259–5274.
- [121] Y.-M. Shao, X. Ma, P. Paira, A. Tan, D.-R. Herr, K.L. Lim, et al., Discovery of indolylpiperazinylpyrimidines with dual-target profiles at adenosine A2A and dopamine D2 receptors for Parkinson's disease treatment, *PLoS One* 13 (1) (2018), e0188212-e.
- [122] L. Xuhan, Y. Kai, V.V. Herman, P.I. Adriaan, J.P.Vw Gerard, An Exploration Strategy Improves the Diversity of De Novo Ligands Using Deep Reinforcement Learning – a Case for the Adenosine A2A Receptor, 2018.
- [123] S. Tian, X. Wang, L. Li, X. Zhang, Y. Li, F. Zhu, et al., Discovery of Novel and Selective Adenosine A2A Receptor Antagonists for Treating Parkinson's Disease through Comparative Structure-Based Virtual Screening, *J. Chem. Inf. Model.* 57 (6) (2017) 1474–1487.
- [124] G.C. Jayson, R. Kerbel, L.M. Ellis, A.L. Harris, Antiangiogenic therapy in oncology: current status and future directions, *Lancet.* 388 (10043) (2016) 518–529.
- [125] T. Voron, O. Colussi, E. Marcheteau, S. Pernot, M. Nizard, A.L. Pointet, et al., VEGF-A modulates expression of inhibitory checkpoints on CD8+ T cells in tumors, *J. Exp. Med.* 212 (2) (2015) 139–148.
- [126] T. Voron, E. Marcheteau, S. Pernot, O. Colussi, E. Tartour, J. Taieb, et al., Control of the immune response by pro-angiogenic factors, *Front. Oncol.* 4 (2014) 70.
- [127] M. Terme, S. Pernot, E. Marcheteau, F. Sandoval, N. Benhamouda, O. Colussi, et al., VEGFA-VEGFR pathway blockade inhibits tumor-induced regulatory T-cell proliferation in colorectal cancer, *Cancer Res.* 73 (2) (2013) 539–549.
- [128] S. Dakshanamurthy, N.T. Issa, S. Assefnia, A. Seshasayee, O.J. Peters, S. Madhavan, et al., Predicting new indications for approved drugs using a proteochemometric method, *J. Med. Chem.* 55 (15) (2012) 6832–6848.
- [129] D. Kang, X. Pang, W. Lian, L. Xu, J. Wang, H. Jia, et al., Discovery of VEGFR2 inhibitors by integrating naïve Bayesian classification, molecular docking and drug screening approaches, *RSC Adv.* 8 (10) (2018) 5286–5297.
- [130] H. Tsukamoto, K. Fujieda, S. Senju, T. Ikeda, H. Oshiumi, Y. Nishimura, Immune-suppressive effects of interleukin-6 on T-cell-mediated anti-tumor immunity, *Cancer Sci.* 109 (3) (2018) 523–530.
- [131] T.A. Mace, R. Shakya, J.R. Pitarresi, B. Swanson, C.W. McQuinn, S. Loftus, et al., IL-6 and PD-L1 antibody blockade combination therapy reduces tumour progression in murine models of pancreatic cancer, *Gut* 67 (2) (2018) 320–332.
- [132] X. Chen, S. Fu, J. Tian, Y. Cao, C. Li, J. Lin, Abstract 864: repurposing FDA-approved drug bazedoxifene as a novel inhibitor of IL-6 signaling for triple-negative breast cancer, *Cancer Res.* 78 (13 Supplement) (2018), 864-.
- [133] H.-W. Han, S. Hahn, H.Y. Jeong, J.-H. Jee, M.-O. Nam, H.K. Kim, et al., LINC1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent, *Sci. Rep.* 8 (1) (2018) 14969.
- [134] C.D. Mills, L.L. Lenz, R.A. Harris, A breakthrough: macrophage-directed Cancer immunotherapy, *Cancer Res.* 76 (3) (2016) 513–516.
- [135] N. Pairet, S. Mang, G. Fois, M. Keck, M. Kuhnback, J. Gindele, et al., TRPV4 inhibition attenuates stretch-induced inflammatory cellular responses and lung barrier dysfunction during mechanical ventilation, *PLoS One* 13 (4) (2018), e0196055.
- [136] S. Tomar, E.E. Zumbrun, M. Nagarkatti, P.S. Nagarkatti, Protective role of cannabinoid receptor 2 activation in galactosamine/lipopolysaccharide-induced acute liver failure through regulation of macrophage polarization and microRNAs, *J. Pharmacol. Exp. Ther.* 353 (2) (2015) 369–379.
- [137] D. Kong, C. Zhou, H. Guo, W. Wang, J. Qiu, X. Liu, et al., Praziquantel targets M1 macrophages and ameliorates splenomegaly in chronic schistosomiasis, *Antimicrob. Agents Chemother.* 62 (1) (2018).
- [138] E. Bok, Y.C. Chung, K.-S. Kim, H.H. Baik, W.-H. Shin, B.K. Jin, Modulation of M1/M2 polarization by capsaicin contributes to the survival of dopaminergic neurons in the lipopolysaccharide-lesioned substantia nigra in vivo, *Exp. Mol. Med.* 50 (7) (2018), 76-.
- [139] P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR Comb. Sci.* 26 (5) (2007) 694–701.

- [140] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, Salakhutdinov R. Dropout, A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* 15 (1) (2014) 1929–1958.
- [141] C. Angermueller, T. Parnamaa, L. Parts, O. Stegle, Deep learning for computational biology, *Mol. Syst. Biol.* 12 (7) (2016) 878.
- [142] M. Sastry, J.F. Lowrie, S.L. Dixon, W. Sherman, Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments, *J. Chem. Inf. Model.* 50 (5) (2010) 771–784.
- [143] D. Rogers, M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.* 50 (5) (2010) 742–754.
- [144] Z. Deng, C. Chuaqui, J. Singh, Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein–Ligand binding interactions, *J. Med. Chem.* 47 (2) (2004) 337–344.
- [145] C. Da, D. Kireev, Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study, *J. Chem. Inf. Model.* 54 (9) (2014) 2555–2561.
- [146] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data* 3 (2016), 160018.
- [147] S.-A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A. Lister, et al., FAIRsharing as a community approach to standards, repositories and policies, *Nat. Biotechnol.* 37 (4) (2019) 358–367.
- [148] V. Stathias, A. Koletti, D. Vidović, D.J. Cooper, K.M. Jagodnik, R. Terry, et al., Sustainable data and metadata management at the BD2K-LINCS data coordination and integration center, *Sci. Data* 5 (2018), 180117.
- [149] S. Abeyruwan, U.D. Vempati, H. Küçük-McGinty, U. Visser, A. Koletti, A. Mir, et al., Evolving BioAssay Ontology (BAO): modularization, integration and applications, *J. Biomed. Semantics* 5 (1) (2014) S5.
- [150] U.D. Vempati, M.J. Przydzial, C. Chung, S. Abeyruwan, A. Mir, K. Sakurai, et al., Formalization, annotation and analysis of diverse drug and probe screening assay datasets using the BioAssay ontology (BAO), *PLoS One* 7 (11) (2012), e49198.
- [151] U.D. Vempati, S.C. Schurer, et al., Development and applications of the bioassay ontology (BAO) to describe and categorize High-throughput assays, in: G. S. Sittampalam, N.P. Coussens, K. Brimacombe, A. Grossman, M. Arkin, D. Auld (Eds.), *Assay Guidance Manual*, 2004. Bethesda (MD).
- [152] Y. Lin, S. Mehta, H. Kucuk-McGinty, J.P. Turner, D. Vidovic, M. Forlin, et al., Drug target ontology to classify and integrate drug discovery data, *J. Biomed. Semantics* 8 (1) (2017) 50.
- [153] N.J. Edwards, M. Oberti, R.R. Thangudu, S. Cai, P.B. McGarvey, S. Jacob, et al., The CPTAC data portal: a resource for Cancer proteomics research, *J. Proteome Res.* 14 (6) (2015) 2707–2713.
- [154] E.G. Armitage, M. Ciborowski, Applications of metabolomics in Cancer studies, *Adv. Exp. Med. Biol.* 965 (2017) 209–234.
- [155] C.H. Johnson, M.E. Spilker, L. Goetz, S.N. Peterson, G. Siuzdak, Metabolite and microbiome interplay in Cancer immunotherapy, *Cancer Res.* 76 (21) (2016) 6146–6152.
- [156] T. Baslan, J. Hicks, Unravelling biology and shifting paradigms in cancer with single-cell sequencing, *Nat. Rev. Cancer* 17 (9) (2017) 557–569.
- [157] S. Sarntinvijai, Y. Lin, Z. Xiang, T.F. Meehan, A.D. Diehl, U.D. Vempati, et al., CLO: the cell line ontology, *J. Biomed. Semantics* 5 (2014) 37.
- [158] U. Visser, S. Abeyruwan, U. Vempati, R.P. Smith, V. Lemmon, S.C. Schurer, BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results, *BMC Bioinformatics* 12 (2011) 257.