



Review article



Small molecule-mediated targeting of microRNAs for drug discovery: Experiments, computational techniques, and disease implications

Jianfeng Sun^{a,1,***}, Miaoer Xu^{b,1}, Jinlong Ru^c, Anna James-Bott^a, Dapeng Xiong^{d,e}, Xia Wang^{f,**}, Adam P. Cribbs^{a,*}

^a Botnar Research Centre, Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, UK

^b Department of Biology, Emory University, Atlanta, GA, 30322, USA

^c Chair of Prevention of Microbial Diseases, School of Life Sciences Weihenstephan, Technical University of Munich, Freising, 85354, Germany

^d Department of Computational Biology, Cornell University, Ithaca, NY, 14853, USA

^e Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY, 14853, USA

^f College of Animal Science and Technology, Northwest A&F University, Yangling, 712100, China

ARTICLE INFO

Keywords:

microRNA targets
Drug discovery
Molecular interactions
Disease implications
Deep learning

ABSTRACT

Small molecules have been providing medical breakthroughs for human diseases for more than a century. Recently, identifying small molecule inhibitors that target microRNAs (miRNAs) has gained importance, despite the challenges posed by labour-intensive screening experiments and the significant efforts required for medicinal chemistry optimization. Numerous experimentally-verified cases have demonstrated the potential of miRNA-targeted small molecule inhibitors for disease treatment. This new approach is grounded in their post-transcriptional regulation of the expression of disease-associated genes. Reversing dysregulated gene expression using this mechanism may help control dysfunctional pathways. Furthermore, the ongoing improvement of algorithms has allowed for the integration of computational strategies built on top of laboratory-based data, facilitating a more precise and rational design and discovery of lead compounds. To complement the use of extensive pharmacogenomics data in prioritising potential drugs, our previous work introduced a computational approach based on only molecular sequences. Moreover, various computational tools for predicting molecular interactions in biological networks using similarity-based inference techniques have been accumulated in established studies. However, there are a limited number of comprehensive reviews covering both computational and experimental drug discovery processes. In this review, we outline a cohesive overview of both biological and computational applications in miRNA-targeted drug discovery, along with their disease implications and clinical significance. Finally, utilizing drug-target interaction (DTIs) data from DrugBank, we showcase the effectiveness of deep learning for obtaining the physicochemical characterization of DTIs.

1. Introduction

Small molecules have long been recognized as a cornerstone in drug development for treating a wide variety of human diseases [1,2]. As of January 2023, the DrugBank database (current version: 5.1.10) has compiled 2736 small molecule drugs that are approved by the Food and Drug Administration (FDA), with over 4500 in preclinical or animal testing stages [3]. Recently, there has been a surge in interest in discovering small molecule inhibitors that specifically target noncoding

RNAs (ncRNAs) [4,5].

Protein-coding genes make up about 1.5% of the human genome [6,7], of which only 3–4% is considered druggable targets [8]. The vast majority of undruggable proteins pose challenges for disease treatment, prompting researchers to explore targeting ncRNAs, especially, to address the issues on dysregulated post-transcriptional regulation [9,10]. ncRNAs are a class of RNA molecules that do not encode proteins, constituting around 70% of the eukaryote transcriptome [8]. Taking into account regions like introns and 3' and 5' untranslated regions, the

* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: jianfeng.sun@ndorms.ox.ac.uk (J. Sun), xiawang@nwfufu.edu.cn (X. Wang), adam.cribbs@ndorms.ox.ac.uk (A.P. Cribbs).

¹ These authors wish to be known as equally contributed.

proportion of non-coding RNA molecules can encompass approximately 95–98.5% of the total RNA proportion in a cell [7,11,12]. The extensive presence of these RNA molecules, combined with their important roles in gene regulation, makes them attractive drug targets [13]. A major class of ncRNAs is microRNAs (miRNAs), which are crucial for regulating cellular activities [14]. Small molecule-mediated targeting of miRNAs can occur before and after miRNA maturation. More recently, research by Childs-Disney et al. has advanced our understanding of the small molecule-mediated targeting of miRNAs by identifying inhibitors that bind to the tertiary folded structures of miRNAs [15]. Subsequently, several other molecules have been identified that work using a similar mechanism [8,16–18], leading to a deeper understanding of small-molecule RNA binding. This understanding has been particularly useful in elucidating how stem-loop hairpin secondary structures are folded into tertiary structures, resulting in the formation of pockets or bulges (such as those found near the Dicer processing site [17]) that aid in small molecule interactions. As a result of this enhanced understanding of small molecule-mediated targeting of miRNAs at the molecular level, researchers have been able to pursue computational rational design for lead compounds and develop algorithms to predict small molecule-miRNA interactions.

Computational approaches for small molecule drug discovery can be broadly categorized into two groups from a structural perspective, based on whether they leverage information about the structure-activity relationships of small molecules [19] and/or the structural details of miRNA targets. In addition to establishing connections between diseases and small molecules, approaches that incorporate such information can also partly or fully facilitate lead compound design [20]. For instance, Inforna has pioneered a novel strategy for discovering small molecule drugs targeting oncogenic miRNA targets by mining from a large dataset of RNA-motif small-molecule interaction pairs, sourced from a microarray-screened library [21]. Specifically, it uses predicted structure annotations of a miRNA of interest to probe highly similar RNA motifs that interact with small molecules. Then, through an algorithmic scoring scheme, the potential small molecules are prioritized for producing lead compounds that are mostly like to treat cancer [22]. On the other hand, the second category often utilizes the information about molecular sequences or simply drug-target as well as target-disease relationships for drug screening [23,24], which can be achieved by using a variety of similarity-based inference methods [25] or machine learning methods [26]. Recently, we developed deep learning approaches to predict the specific regulation types between a small molecule and a miRNA, relying solely on molecular sequence information [27]. Subsequently, we implemented a drug repositioning strategy that connects a drug and a disease using calculated connectivity scores through the overlapped disease and drug relationships with miRNAs (as in Ref. [28]). These connectivity scores were introduced to facilitate drug repositioning based on distinct upregulation and downregulation information [29], which was obtained through statistical tests using gene/transcriptional expression data from large-scale pharmacogenomic studies [30,31]. Nonetheless, the scarcity of miRNA expression data associated with numerous small molecules makes it difficult to address using existing techniques, as generating such data requires considerable experimental effort. In this context, computational strategies offer an alternative solution to readily generate information about miRNA regulation.

In this review, we attempt to summarize the small molecule-mediated targeting strategies for the discovery of drugs for both protein and miRNA targets. We cover the functional importance of miRNAs as small-molecule targets, disease implications, clinical applications, and miRNA pharmacogenomics. We also benchmark three key prediction problems in early drug discovery: predicting small molecule-miRNA (SM-miR) associations [32], SM-miR regulation types [27], and drug-target interactions (DTIs) [33]. To assess the current status of the DTI prediction based on sequences alone, we analysed the DrugBank database using deep learning methods, achieving an area under the

curve (AUC) of approximately 95% for independent test DTIs and 75% for those involving novel drugs. Recognizing the transformative potential of artificial intelligence, we additionally provide an overview of key deep learning methodologies in DTI prediction [34,35]. Our review complements existing literature by integrating experimental and computational aspects of drug discovery.

2. miRNA dysregulation in diseases

Due to their critical role in gene expression regulation, miRNAs have been implicated in a wide spectrum of diseases [14,36]. For example, the overexpression of miR-17–92, a widely studied miRNA cluster [37], is found to result in inhibiting proximal epithelial cell differentiation [38] and polycystic kidney disease (PKD) [39]. Regulus Therapeutics has developed a first-in-class oligonucleotide-based drug RGLS4326 to treat PKD by targeting miR-17, a member of the miR-17–92 family [37]. Its clinical trials are currently underway [40,41]. Also, long noncoding RNAs (lncRNAs), another major component of ncRNAs (>100,000 human lncRNAs [42]), are highly associated with a large number of human diseases [43,44]. For example, the lncRNADisease database (v2.0) has currently compiled 10,002 lncRNA-disease associations that are detected in humans, which account for 94.7% of all the associations [45]. Dysregulated expression levels of multiple miRNAs are often observed in a single disease. For example, Yanaihara et al. conducted a correlation analysis of miRNA expression profiles in patients with lung adenocarcinoma and discovered the aberrant expression of several miRNAs, such as the upregulation of miR-155, the downregulation of let-7a-2, and the downregulation of miR-145. The expression of these miRNAs was found to be correlated with the patient survival time [46]. miRNAs can also work in conjunction with other types of molecules to exert and amplify oncogenic effects in diseases. For example, the Myc oncogene (encoding a transcription factor in charge of transcription events of ~15% genes [47]) and miR-155 have been seen to be co-overexpressed to promote B-cell lymphomas [48]. These findings imply that the complex interplay of a combination of dysregulated miRNAs may contribute to the severity of certain diseases.

3. Small molecules targeting miRNAs for cancer therapeutics

Physical interactions of multiple kinds of biological molecules with miRNAs of interest provide the potential to alter the functions of the miRNAs, which has made it possible to regulate their controlled genes that are abnormally expressed in human diseases. This has opened up extensive possibilities for developing miRNA targeting strategies using small molecules [4], miRNA sponges [49], or antisense oligonucleotides [50]. The efficacy of using oligonucleotides and small molecules is discussed in detail as in Refs. [17,21], which indicate that small molecules possess several advantages over oligonucleotides, such as greater selectivity for inhibition and increased cell permeability under specific conditions. Our review is centred on the targeting of miRNAs by small molecules. Rather than directly inhibiting disease-related gene targets, small molecules achieve therapeutic effects by regulating the transcriptional expression of miRNAs, which subsequently affects the expression of malfunctioning gene targets indirectly [4]. Small molecule-mediated regulation of miRNAs can occur at many stages over the course of miRNA biogenesis (see Fig. 1). Specifically, small molecules can achieve therapeutic effects not only by directly binding to mature miRNAs but also by binding to the functional sites (such as DROSHA processing sites) in miRNA precursors where mature miRNAs are located, inhibiting the biogenesis of miRNAs, as suggested by Childs-Disney et al. [15]. Aside from targeting the functional sites, Li and Rana described two additional methods for regulating miRNA expression: blocking the assembly of pri-miRNAs and obstructing the formation of the miRNA-induced silencing complex (miRISC) [4], which affect the miRNA maturation and mRNA degradation processes, respectively [51]. Similar to disrupting the miRISC, small molecules

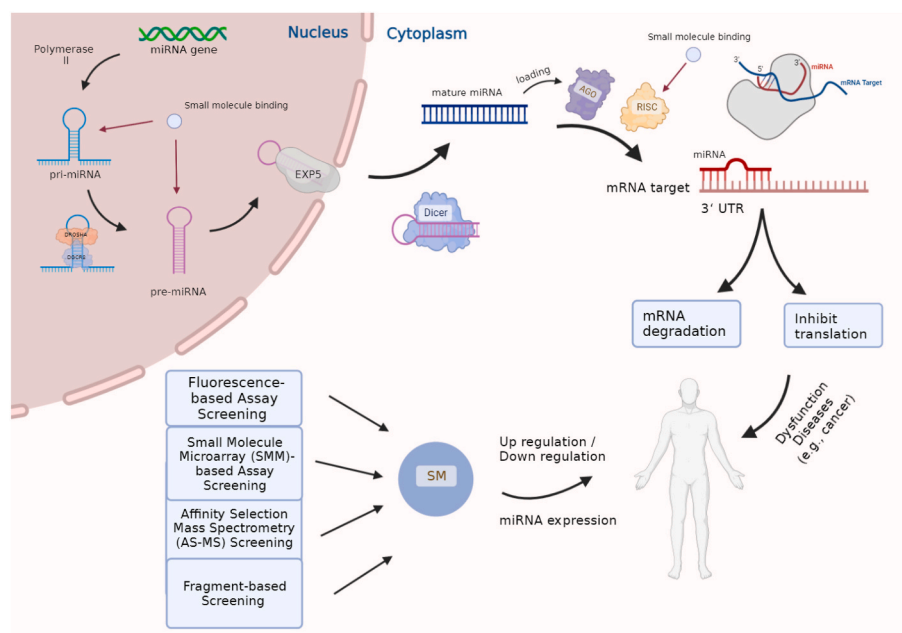


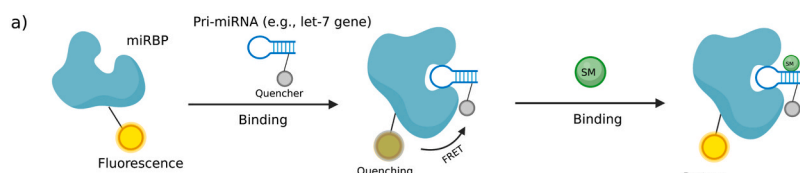
Fig. 1. Small molecule therapeutics targeting the biogenesis of miRNAs. In the cell nucleus, miRNA genes are first transcribed by RNA polymerase II into primary transcripts (pri-miRNAs) whose mono-cistronic, polycistronic, and/or intronic regions (see a comprehensive review [249]) possess the characteristic hairpin secondary structures [51] in which mature miRNA sequences are located [250]. After being recognized by DGCR8 (i.e., DiGeorge critical region 8, the RNA-binding protein) [251] and Drosha (i.e., the RNase III enzyme) [252], these hairpin structures are then cleaved to produce the precursor miRNAs, known as pre-miRNAs [253]. Subsequently, the pre-miRNAs are transported by exportin-5 [254] to the cytoplasm, in which they are converted into mature miRNAs through the Dicer-mediated processing mechanism [245,255]. The mature miRNAs culminate in being loaded with the Argonaute (AGO) protein into a bioactive miRNA-induced silencing complex (miRISC). The loaded miRNAs will guide miRISCs to initiate its binding to the complementary sequences in the 3'untranslated region (UTR) of their mRNA targets [256]. The result of miRNA binding is the silencing of the target gene [58], which occurs in a phased manner over time [256]. During the early phase, the target gene's translation is inhibited [249], while in the later phase, mRNAs transcribed by the affected genes are degraded through mechanisms such as mRNA deadenylation [257]. This phased

progress gives miRNAs two distinct roles: the inhibition of gene translation and the degradation of mRNAs [258]. The dysregulation of the expression of the miRNAs will cause the onset of their regulated disease pathways, while the small molecule identified as being helpful for inhibition can be used to rectify the dysregulation of the miRNA expression for disease treatment. For example, this can be achieved by targeting pri-miRNAs and pre-miRNAs within the nucleus and miRISCs in the cytoplasm.

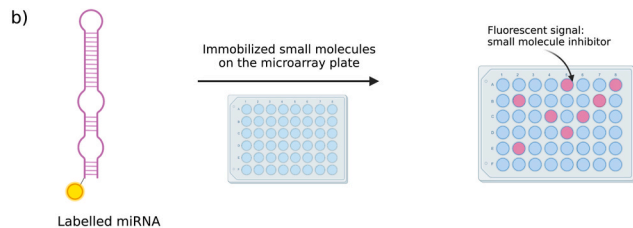
could target other miRNA-binding proteins (miRBPs) to modulate miRNA expression, such as LIN28 [52], Toll-like receptors [53], and non-canonical bacterial RBPs [16]. The mechanism of small molecule

therapeutics is self-explanatory, as they can induce a change in gene expression profiles in diseases by targeting miRNAs, irrespective of whether the altered miRNA expression arises from mRNA degradation

Fluorescence-based Assay Screening



Small Molecule Microarray (SMM)-based Assay Screening



Affinity Selection Mass Spectrometry (AS-MS) Screening

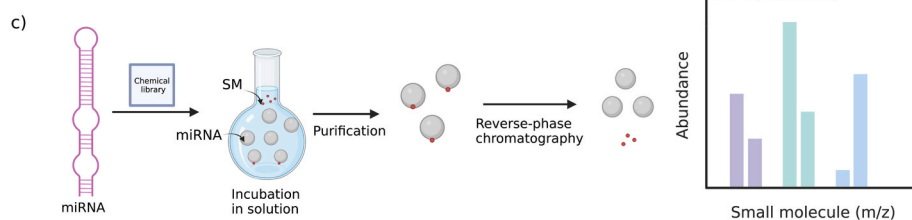


Fig. 2. Illustration of chemical approaches for target-centric screening of small molecules in a high-throughput manner. (a) Fluorescence resonance energy transfer (FRET)-based screening. The experiment parameters of this approach are set in a way that a miRNA-binding protein (miRBP, e.g., Lin28A and Lin28B [16]) is tagged with a fluorescence indicator and a quencher is attached to a primary miRNA (pri-miRNA) encoded by the let-7 gene. The binding of the miRBP to the pri-miRNA will induce quenching of the fluorescence. A small molecule inhibitor of the pri-miRNA will restore the fluorescence through disruption of the binding [16,58]. (b) small molecule microarray (SMM)-based assay screening. Small molecules are immobilized on the microarray plate to await the binding of a labelled miRNA [15,59]. Then, high-fidelity signals are shown if the binding occurs. (c) Affinity selection mass spectrometry (AS-MS) screening. Comparatively, the AS-MS-based approach uses label-free targets to screen small molecule inhibitors. The miRNA target is incubated for binding to small molecules. Their complexes are first purified using the chromatography technique and then small molecules are dissociated from the miRNA target using the reverse-phase chromatography technique. Finally, the abundance of different small molecules that can bind to the target is generated using AS-MS. This review collects abundant case studies by using these varying HTS techniques [259].

or miRNA assembly processes.

4. Experimental determination of interactions between small molecules and miRNAs

The increasing number of experimentally-evidenced cases from the RNA-targeted drug discovery field demonstrates the utility of small molecule inhibitors in targeting specific miRNAs (SMIRs) for disease treatment [54]. It has been estimated that the quantity of RNA targets is roughly of at least two orders of magnitude as that of protein targets [55]. In contrast to the growing number of experimentally resolved protein structures in the protein data bank (PDB), the accumulation of RNA-only structures has remained stagnant, with an annual of only 70–100 annually deposited structures over the past decade (~1% compared to proteins [56]). Moreover, it is likely that miRNAs only make up a small proportion of these structures.

In addition to high-fidelity structure models, there are several high-throughput screening (HTS) techniques available to precisely and promptly determine SMIRs from a large compound library, to identify lead compounds targeting specific miRNAs involved in human disease pathogenesis [57]. These established techniques mainly include, fluorescence-based assay screening, small molecule microarray (SMM)-based assay screening, fragment-based screening, and affinity selection mass spectrometry (AS-MS) screening [4,8,15,17,51]. These methods are destined for high-throughput screening of large small molecule libraries against one or more specific miRNA target(s), as shown in Fig. 2. Fluorescence-based assay screening approaches possess several derivatives depending on detecting the intensity from fluorescence molecules collectively or individually [58]. In particular, an example of such a derivative is the fluorescence resonance energy transfer (FRET)-based screening approach, which relies on the principle that the fluorescence quenching will not occur if a small molecule inhibitor successfully targets a quencher-attached miRNA that has physically bound to a fluorescence-labelled molecule [16,58]. We only summarized FRET-based screening approach in this review. Small molecule inhibitors are identified by the SMM-based approach in a way that high-fidelity signals are shown on the microarray, in which a library of small molecules are immobilized to attempt to bind to labelled miRNA targets [15,59]. The fragment-based approach leverages a library of small molecules of low molecular weight as fragments to vet the possibilities of them binding to miRNA targets. Then, these possibilities are translated into signals by using, for example, NMR spectroscopy, which are partitioned into two groups to be indicative of bound and unbound fragments, respectively (as depicted in Ref. [60]). Comparatively, the AS-MS-based approach uses label-free targets to screen small molecule inhibitors. Using automated ligand identification system (ALIS), a type of the AS-MS-based approach, binding small molecule ligands are first dissociated from chromatography-purified ligand-target complexes after incubation and then examined from mass spectrometry signals. In addition, there are also a few other approaches for this purpose, such as phenotype-based screening [20], DNA-encoded compound library-based screening [15], and pharmacological validation screening [61]. It is worth mentioning that the phenotype-based approach can work by overlooking the reliance on the knowledge about disease-associated targets and shortlists small molecules that are able to counteract the pathogenic effect of a disease-associated phenotype [57]. The main characteristics of the approaches are summarized in Table 1.

5. Clinical applications of ncRNA drugs or targets

Despite the tremendous efforts made to study small molecule inhibitors targeting ncRNAs, clinical applications are still in their infancy [40]. However, miRNAs play an important role in being targeted by drugs or targeting other disease-associated genes [62]. According to the statistics presented by Zhang et al. [63], there are tens of siRNA (as drugs)- and miRNA (as targets)-related therapeutics that have been

Table 1

Summary of important characteristics of several mainstream HTS methods.

HTS Method	Material	Note
Fluorescence resonance energy transfer (FRET)-based screening	Quencher and fluorescence	The binding of a small molecule inhibitor to the pri-miRNA prevents quenching.
Small molecule microarray (SMM)-based assay screening	Immobilized small molecules, microarray plates, fluorescence microscopy	Fluorescence intensities that appear and deepen in a pore on the microarray indicates the binding of a small molecule inhibitor to the miRNA target.
Affinity selection mass spectrometry (AS-MS) screening	Chemical libraries, chromatography, reverse-phase chromatography, mass spectrometry	Purification of compound-binding miRNAs as many as possible is necessary for accurate quantification.
Fragment-based screening	Small molecule libraries, NMR spectroscopy	A library of small molecules of low molecular weight as fragments to vet their possibilities of binding to miRNA targets.

entered into different phases of clinical trials. For example, RG-101, developed by Regulus Therapeutics, is an GalNAc-conjugated anti-miRNA oligonucleotide sequence to target miRNA-122 against hepatitis C virus infection [40]. However, small molecule inhibitors of miRNAs remain to be tested out only in the lab. For example, using a small-molecule microarray platform called AbsorbArray, a couple of topoisomerase inhibitors were shown to bind to the A bulge in the Dicer site of pre-miR-21 for inhibition of miR-21 expression, which affected the expression of their regulated genes in triple negative breast cancer [64].

In addition, lipid nanoparticle (LNP) drug delivery systems loading with siRNA-based drugs (trade name: Onpatro) has been approved by FDA to treat polyneuropathies [65]. Like the low success rate of discovering drugs for targeting other disease-related agents (i.e., proteins), the commercial development of miRNA therapeutics is confronted with a high failure rate, with a large majority of the therapeutics suspended over the course of drug discovery. But the analysis from Ref. [63] also demonstrates high rates of ncRNA therapeutics (40% for both miRNAs and siRNAs) that are being subjected into phase II clinical trials or more. Nevertheless, generating miRNA therapeutics for disease treatment is still of prime interest to a number of pharmaceutical industries, including miRagen Therapeutics and Regulus Therapeutics [66,67].

6. Experimental determination of drug-target interactions

In this review we refer to the interactions between drugs and protein targets as drug-target interactions (DTIs). Mass spectrometry (MS)-based chemical proteomics approaches, including activity-based protein profiling (ABPP) [68] and compound-centric chemical proteomics (CCCP) [69], have been introduced as a powerful tool in high-throughput to yield a large number of interactions between small molecule drugs and protein targets [70,71]. In addition, the evidence for biophysical binding and structural details can also be detected through functionally active protein-compound complexes by the nuclear magnetic resonance (NMR) spectroscopy [72]. The modulation of protein functions is achieved by targeting the protein-protein interaction (PPI) interfaces by chemical compounds [73–75]. As highlighted in Refs. [76,77], NMR excels at capturing the fine-grained local chemical environment/structure compared to cryo-electron microscopy (cryo-EM) [78] as well as X-ray crystallography [79], which makes it more capable of determining DTIs. It is certain that there exist alternative methods for biophysical interaction detection (see this review [80]). There is no doubt that membrane proteins constitute a predominant proportion of

all targets [81,82], since they are estimated to involve the transduction of around 85% cell signals [83]. Based on the most recent DrugBank database (version: 5.1.9) [84], we calculate around 50% of all targets as transmembrane proteins, which are targeted by FDA-approved drugs.

7. Connectivity scores for drug repositioning

The treatment of cells with drugs can induce changes in gene expression levels, leading to a variety of gene expression signatures (a gene or a group of genes, cf. deregulated genes [85] and drug perturbation signatures [86]) that are differentially expressed as a result of cellular responses to this treatment [87]. Such drug-induced gene expression signatures are conducive to decipher a certain gene expression pattern(s) occurring in a disease and further favour the drug development for the disease [29]. This has given rise to large-scale clinical pharmacogenomic studies [88,89], which yield massive datasets of gene expression matrices induced by a large quantity of drugs, termed drug perturbation datasets [90]. In comparison, there are also drug sensitivity datasets for screening cancer cell lines (see Refs. [31,91,92]). Note that main approaches for miRNA expression profile measurement comprise qRT-PCR, microarrays, and RNA-seq (for details, see Ref. [93]). These pharmacogenomic datasets can be experimentally generated through many cancer models [94], such as patient-derived cancer cell lines [95], organoids [96], and xenografts (PDXs) [97], which serve as valuable resources to predict drug responses and identify biomarkers. We focus on the drug perturbation datasets in the following parts. Despite rapid growth in the number of online resources that are powerful for studying drug perturbation behaviours (e.g., CMAP [29] and a scale-up version of CMAP, L1000 [98], see also a review [30]), there might exist some concerns in terms of the size of the data as well as the quality of the data, as pointed out by Sharifi-Noghabi et al. [91] and Keenan et al. [99], respectively. Nevertheless, by exploiting drug perturbation datasets in conjunction with disease-gene datasets, the associations between drugs and diseases can be established.

Connectivity scores have been developed to infer such associations between drugs and diseases. The recently developed recommended connectivity-map scoring method (RCSM) package serves as a repertoire of connectivity scores [28], which contains, for example, the gene set enrichment analysis (GSEA) method and the reverse gene expression score. In the RCSM package, the KS module ranks the gene expression signature of a drug based on the log fold change of gene expression in a bidirectional fashion. In short, this process initiates the classification of the downregulated and upregulated genes of a gene expression signature, and moves to rank these genes within each category, which are however retained in the same gene list where the upmost gene is the one with the biggest log fold change value and the downmost gene is the one with the smallest log fold change value. However, differentially expressed genes (DEGs) identified via a threshold of log-fold change values together with the p-value information based on gene expression matrices (especially miRNA expression matrices) may not easily be accessible due to many ad hoc experimental efforts. Rather than using the information about DEGs derived from expression matrices, the drug repositioning process presented on the DeepsmirUD-Web website exploits a new miRNA expression regulation profile predicted by DeepsmirUD (a recently released deep learning implementation) [27], which we referred to as small molecule-mediated regulatory effects on miRNA expression in the DeepsmirUD work. DeepsmirUD performs the inference of the small molecule-mediated downregulation and upregulation profiles of miRNA expression based solely on molecular sequences of miRNAs and small molecules without relying on an expression matrix specific to a list of drugs in the treatment of disease in a certain cell state. Similarly, this miRNA expression dysregulation profile allows the classification of downregulated and upregulated miRNAs. But differently, it cannot be used to indicate DEGs based on statistical tests because the predicted value for each miRNA represents how likely the miRNA expression is downregulated or upregulated by a query small molecule.

Correspondingly, the most top-ranked miRNA represents the one that is most strongly inferred as being upregulated by the query small molecule. We stress this difference in terms of the utilization of statistical tests for examination of the extent of the differential expression genes profile of a gene signature, which makes this scheme a trade-off for drug repositioning based on molecular sequences and connectivity scores. Nevertheless, the DeepsmirUD-predicted downregulation and upregulation profiles make it possible to infer drug-disease associations by virtue of only sequences. These results may be useful to some degree, insofar as we presented a few successfully inferred examples in the DeepsmirUD work [27]. The comparison between the standard scheme and our piloted scheme is schematically illustrated in Fig. 3. We detail the formulation of this drug-disease association process involving DeepsmirUD predictions as follows.

The crux of this idea for drug discovery is the leveraging of statistical tests of genes that are overlapped between the drug-induced gene signature and a list of disease-associated genes [100,101]. One of the earliest connectivity scoring schemes is a type of gene set enrichment score (GSES) calculated based on the non-parametric Kolmogorov-Smirnov (KS) test method [29,102,103], which is integrated into the CMAP 1.0 database [28]. We dubbed this method GSES-KS in this review. Because of its use without relying on an ordered list of disease-associated genes, we have previously introduced this GSES-KS method into the DeepsmirUD work to discover drugs using an unordered list of disease-associated miRNAs from miRCancer [104]. Two KS statistics are first constructed as

$$a = \max_{0 \leq i \leq N_{miR}} \left(\frac{i}{N_{miR}} - \frac{p}{L_{SM-miR}} \right)$$

$$b = \max_{0 \leq i \leq N_{miR}} \left(\frac{p}{L_{SM-miR}} - \frac{i-1}{N_{miR}} \right)$$

where the miRNA expression signature of drug *A* is denoted as L_{SM-miR} that contains N_{sm-miR} miRNAs and disease-associated miRNA list is denoted as L_{dz-miR} that contains N_{dz-miR} miRNAs. The number of miRNAs that are overlapped between L_{SM-miR} and L_{dz-miR} is N_{miR} . p represents the position i^{th} of a gene in the ranked L_{SM-miR} miRNA expression signature in ascending order of the DeepsmirUD-predicted values in both downregulation and upregulation directions.

$$\begin{cases} GSES - KS^+ = a, a > b \\ GSES - KS^- = -b, a < b \end{cases}$$

where the $GSES - KS^+$ score indicates a profile for upregulated miRNAs in a given disease and the $GSES - KS^-$ score indicates a profile for downregulated miRNAs in the disease. Then, the connectivity score $GSES - KS$ is given by

$$GSES - KS = \begin{cases} GSES - KS^+ - GSES - KS^-, c \\ 0, otherwise \end{cases}$$

where c means that both $GSES - KS^+$ and $GSES - KS^-$ are positive or negative values. Finally, the connectivity score is normalized as $(GSES - KS)_{norm}$ ranging from -1 to 1 . It should be noted that the availability of multiple connectivity scores actually allows much flexibility in building the relationships between the DeepsmirUD-predicted miRNA expression signature and the disease-associated miRNAs. In effect, the application of the connectivity scoring schemes or the CMAP databases has begun to extend to a few fields, including identifying active ingredients from traditional Chinese medicine [105,106].

8. Reverse pharmacological profiles of structurally similar small molecules

It is crucial to exercise caution when processing drug-target association/interaction data, as it may significantly impact the quality of

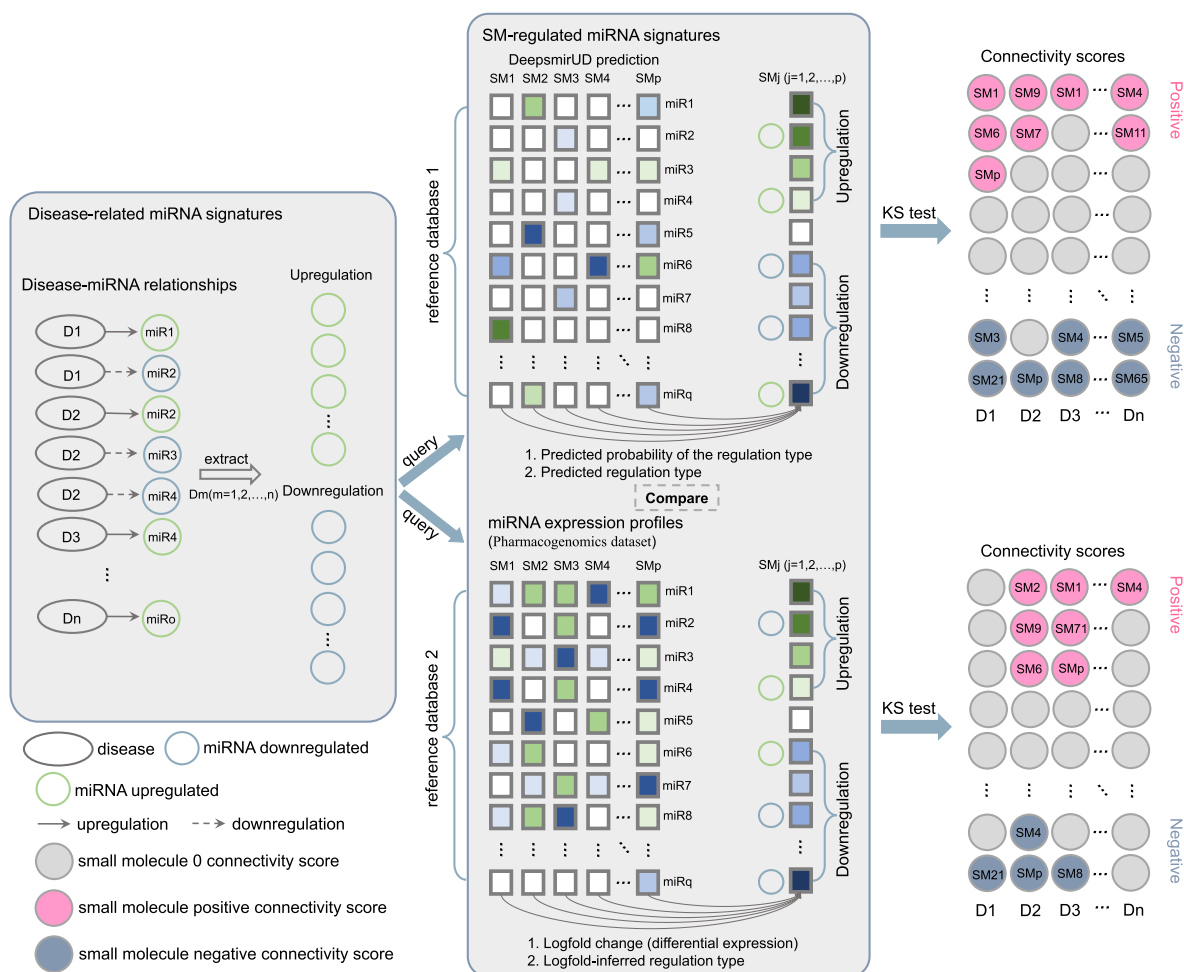


Fig. 3. Drug repositioning using connectivity scores. miRNA signatures of a disease of interest are scanned from a database of disease-miRNA relationships and represented by two groups of miRNAs: upregulation or downregulation. Conventionally, the drug repositioning process then queries a reference database for pattern matching of small molecule-induced miRNA signatures with the disease-related miRNA signatures. Such a reference database is obtained from purpose-built large-scale clinical pharmacogenomic studies and represented by a small molecule-induced gene expression matrix. Within each small molecule, the gene signature is generated and ranked by differential expression analysis. The symbols of the resulting logfold change values indicate whether genes within the small molecule are downregulated or upregulated. The ranks of genes within the small molecules are decided by the logfold change values (see Ref. [28] for explanation). Due to the time-consuming nature of the experiment process, the gene expression profiles cannot be accessed easily. Our piloted scheme allows for generating the regulation effect of each small molecule on the expression of each miRNA through only their biological sequences. In this scheme, DeepSmirUD-predicted regulation profiles are used to replace the small molecule-induced gene expression profiles. In comparison, genes within each small molecule are identified as being downregulated or upregulated by the predicted regulation type while their ranks within the small molecule are decided by the predicted probabilities. The final connectivity scores are generated by the Kolmogorov-Smirnov (KS) statistic test to indicate the druglike potential of small molecules using the two routes, respectively.

computational analysis and prediction. Small changes in the structure of a molecule can lead to large perturbations in its chemical activity, a phenomenon that is well known in molecular design [107–109]. It has been suggested that a slight chemical modification on an antagonist, which binds to the human A3 adenosine receptor (hA³AR), turns it to an agonist [110]. In the hA³AR-agonist model, the agonistic compound induces an intense conformational change of the hA³AR, while in the hA³AR-antagonist model, the antagonistic compound keeps the hA³AR activity in check [110,111]. By comparing the two structures, we find that two compounds derived from their respective models are shown to be highly structurally similar, as evidenced by a Tanimoto coefficient of 0.992. Both agonistic and antagonistic modes of this compound are of pharmaceutical significance to act on the pathogenic mechanisms of different diseases as potential drugs. However, it poses a grand challenge to computational analysis. For instance, if the two triplets of drug-target responses (hA³AR-agonist-activation and hA³AR-antagonist-inhibition) are catalogued into a cohort for a two-label classification problem, the optimization process or the ultimate performance of machine learning models could deteriorate by simultaneously considering the two samples

with extremely similar features (i.e., hA³AR-agonist) but completely different labels (i.e., activation and inhibition). This is because the association types (labels) that are formed biologically are sometimes independent of the sequence or structural similarity between small molecules, miRNAs, or both. Therefore, this could potentially complicate the situation of reducing the redundancy of drug and target data, i.e., particularly whether to remove those similar samples across all categories (activation and inhibition) or simply within each category.

9. Deep learning

9.1. Convolution-based deep learning architectures

Deep learning algorithms are methods that learn representations by automatically extracting features from input data [112]. This enables a diminished involvement of domain experts with the preparation of crafted features, which makes it possible to gain a high degree of accuracy that is equivalent to the utilization of features from a tedious, arduous preparation process [113]. The learning process involves linear

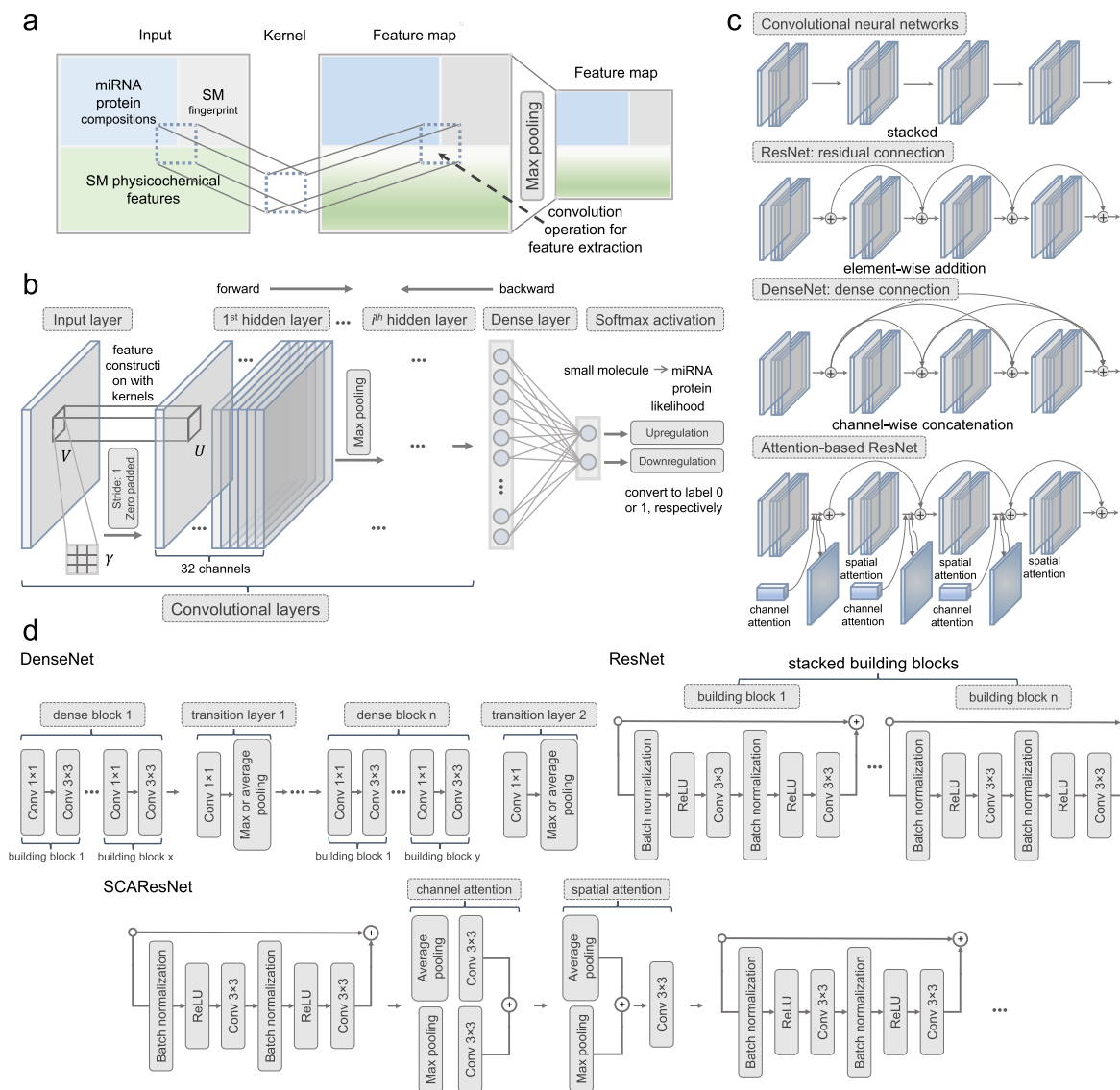


Fig. 4. Deep learning approaches for learning and discovering the pattern of relationships (i.e., regulation type or interaction) between small molecules and miRNAs/proteins. (a) Convolution-based deep learning methods extract small molecule and miRNA features. Using the convolution techniques, their initially concatenated feature representations are progressively fused together in their feature maps at high-order neural layers. For example, the boundaries that initially demarcated by different types of features blur as a result of filtering with convolution and subsampling using max pooling. (b) A typical convolutional neural network (CNN) showing a stack of convolutional neural layers placed with multiple filters/kernels for feature extraction over initial input image-like objects of biological sequence features or feature maps, which, after going through a densely connected layer, are processed into highly condensed values to be indicative of relationships between small molecules and miRNAs/proteins. (c) shows four different ways to connect between every two bundles of convolutional neural layers for a CNN (serially connected), a ResNet (residually connected), a DenseNet (densely connected), and an attention-based ResNet (connected with spatial and channel attention modules). (d) shows the detailed building blocks of DenseNet, ResNet, and SCAResNet.

(e.g., the addition operation) or non-linear (e.g., the ReLU activation function [114]) transformations of input data being fed into neurons from one to another neural layer [115]. To satisfy these needs, deep learning algorithms often have proprietary architectures, which are constructed using a few basic unit layers, including convolutional layers [116], recurrent layers [117], and graphical convolutional layers [118]. As summarized in this review [119], the same or different types of these layers are assembled into convolutional neural networks (CNNs) [120], recurrent neural networks (RNNs) [121], and graphical convolutional neural networks (GCNs) [122] through, for example, parameter sharing settings [123]. For deep neural networks pertaining to convolutional layers, subsampling is inevitably involved in their assembly processes as it can reduce computational costs significantly [124]. Common subsampling strategies include channel-wise or element-wise max-pooling or average-pooling operations [125].

We have found the use of deep learning architectures of convolutional layers connected in multiple ways to be highly effective in predicting small-molecule-mediated regulatory effects on miRNA expression [27] as well as drug-target interactions (given in sections below). Therefore, we mainly confine the deep learning part of this review to the introduction of the convolution-based models. We present a couple of representative architectures, including ResNets [126], DenseNets [127], and SCAResNets [128], that produce a series of stably varying values of evaluation metrics (e.g., AUC or AUCPR) across training epochs than other types of architectures in the two learning tasks above. The main characteristics of the models used in the DeepSmirUD work are summarized in Table 2. The similarity and difference between these models are summarized as follows.

Table 2
Deep learning models adopted in the DeepsmirUD work.

Model	Conv layer number	BB number	Conv 1×1 number	Conv 3×3 number
CNNs	3	1 (1)	–	3
ResNet18	17	8 (2, 2, 2, 2)	–	16
ResNet50	19	16 (3, 4, 6, 3)	32	16
DenseNet41	40	18 (6, 3, 3, 6)	21	18
SCAResNet18	17	8 (2, 2, 2, 2)	–	16

Note: Conv: convolutional. BB: building block. Conv 1×1 is used for setting a bottleneck architecture that can significantly reduce the number of parameters in filters and yet keep the dimension of feature maps unchanged [126]. Each number separated by a comma in brackets represent the number of basic building blocks in each stacked building block structure (see also Fig. 4d).

Table 3
Examples of convolution-based deep learning architectures.

Architecture	Structure within a BB	Connection between BBs	Subsampling	Attention mechanism
CNNs	–	–	max pooling	No
ResNets	conv layers, ReLU, BN	identity shortcut	max/average pooling	No
DenseNets	conv layers	dense connection	max/average pooling	No
SCAResNets	conv layers, ReLU, BN	identity shortcut	max/average pooling	Yes

Note: Conv: convolutional. BB: building block.

- Similarity.** Apart from taking all convolutional layers as its backbone structure, all these models culminate with a fully-connected layer for converting all previously learned features into a possibility for indicating a binary label (e.g., upregulation or downregulation). These models are structured as a connection of building blocks, each alternating between batch normalisation, non-linear activation functions, and/or convolutional layers in a certain order (see Table 3). For example, for the pre-activation design of a building block, batch normalisation is placed first, which is followed by the ReLU activation function and a convolutional layer. All the three types of models utilize average pooling or max pooling for subsampling to reduce the size of feature maps.
- Difference.** The convolution-based deep learning models differ primarily in the way how building blocks connect to each other and the time required for training.
 - Building block connection.** As shown in Fig. 4c, there are four ways to connect between convolutional layers. The raw CNNs are a stack of serially connected convolutional layers without applying additional operations [129]. In a residual neural network, every two convolutional layers are connected in a residual manner as a residual unit in which the final output of it is the sum of the output from the second convolutional layer and the raw input to the first convolutional layer, which is also called the identity shortcut (see Table 3). The invention of ResNets is conceived as a milestone in the development of deep learning algorithms, since this type of network is widely believed to allow for an ultrahigh speed to train an ultradeep CNN with the layer number equal to it. Inspired by the residual connection scheme, Huang et al. designed an architecture of densely connected layers, termed DenseNets, where convolutional blocks, each represented by a few serially connected convolutional layers, are connected in a way that the output of all convolutional blocks is the sum of the output of the last convolutional block and the sum of the output of all the preceding convolutional blocks [130]. The main difference between the two types of connections is that DenseNets allow for reusing the

extracted features from all previous building blocks while ResNets can only use the feature information from its preceding one building block. The performance of DenseNets was tested to be superior to that of ResNets. Another study employed an attention-based mechanism [131] to further enhance the prediction performance using ResNets [192]. The approach incorporated spatial and channel-wise attention (SCA) modules onto the initial output of every residual unit. These SCA modules, together with the raw input to that residual unit, are then added up to the input to the next residual unit. Essentially, the SCA modules reinforce feature extraction from specific regions of feature maps by introducing two extra sets of weights in addition to the weights of the kernels in each residual unit. A more comprehensive explanation can be found in Ref. [132].

- Time and space complexity.** The development of new deep learning architectures for image recognition has been focused on shortening their training time at no cost of prediction accuracy or at even improved accuracy [133]. Since the parameters of those architectures involving convolution operations are sourced solely from filters, the size and number of filters are of direct relevance to the issue on computational costs. Theoretically, the time complexity of executing a single convolution operation in a convolutional layer is kept unchanged if the input and the filter are set as the same in that layer, regardless of whatever types different deep learning architectures are. However, there can be much pronounced difference between deep learning architectures in terms of the space complexity. For example, DenseNets consume much more Ram/GPU memory than ResNets to handle more complex concatenation operations between building blocks. It is reported that compared to ResNets, the use of DenseNets leads to a 1.2-fold increase in the training time [134]. Since SCAResNets are constructed with additional attention modules (i.e., additional convolution) on top of ResNets, they are doomed to require longer time for model training than ResNets.

9.2. Weight determination using the backpropagation algorithm

The above derivatives of CNNs are built by hinging on a basic structure, namely, the convolutional layer in which kernels (also called filters containing weights or parameters) are placed to perform convolutional operations on image objects along horizontally and vertically. We next set out to mathematically formulate this process. To made it clear, we show in Fig. 4b that a kernel between two adjacent layers contains the weight matrix γ that is required to be determined, which can be achieved by the backpropagation algorithm [135,136]. CNNs are feedforward neural networks through which information is propagated along the forward- and backward-pass directions [137,138], as indicated by upper arrows in Fig. 4b. During a forward pass, the information flows from a feature map in layer l to another feature map in layer $l + 1$ [139], which is written as

$$u_{m,n}^{l+1,d} = \sum_{d=0}^D \sum_{p=0}^P \sum_{q=0}^Q \gamma_{p,q}^{l+1,k,d} \times v_{m+p,n+q}^{l,d}$$

where D is said to be the number of channels, and P and Q are the length and the width of a kernel, respectively. $\gamma_{p,q}^{l+1,k,d}$ represents the weight at row p and column q in the weight matrix γ (i.e., the k th kernel) at the d th channel in layer $l + 1$. $v_{m+p,n+q}^{l,d}$ represents the element (i.e., a small molecule or a miRNA feature as in the DeepsmirUD work) at row $m + p$ and column $n + q$ in the feature map V at the d th channel in layer l . $u_{m,n}^{l+1,d}$ represents the element at row m and column n in the output U at the d th channel in layer $l + 1$. It can also be rephrased to its matrix form, such that,

$$U^{l+1,d} = \sum_{d=0}^D \gamma^{l+1,k,d} \otimes V^{l,d}$$

where \otimes represents the convolution operation. Let J be the cost function. During the backward-pass process, in each epoch, the weight $\gamma_{p,q}^{l,k,d}$ at row p and column q in the weight matrix Y at the d th channel in layer $l+1$ is updated by the gradient descent method [140,141].

$$\gamma_{p,q}^{l+1,k,d} = \gamma_{p,q}^{l,k,d} - \alpha \frac{\partial J}{\partial \gamma_{p,q}^{l+1,k,d}}$$

$$\frac{\partial J}{\partial \gamma_{p,q}^{l+1,k,d}} = \sum_{m=0}^{M-P} \sum_{n=0}^{N-Q} \frac{\partial u_{m,n}^{l+1,k}}{\partial \gamma_{p,q}^{l+1,k,d}} \times \frac{\partial J}{\partial u_{m,n}^{l+1,k}}$$

$$\frac{\partial J}{\partial \gamma_{p,q}^{l+1,k,d}} = \sum_{m=0}^{M-P} \sum_{n=0}^{N-Q} \frac{\partial J}{\partial u_{m,n}^{l+1,k}} \times v_{m+p,n+q}^{l,d}$$

$$\frac{\partial J}{\partial \gamma_{p,q}^{l+1,k,d}} = \frac{\partial J}{\partial U^{l+1,k}} \otimes V^{l,d}$$

where M and N represent the length and the width of an input feature map. $\frac{\partial J}{\partial U^{l+1,k}}$ is calculated based on its counterpart in layer $l+2$, as demonstrated in Ref. [141]. The loss, which is the difference between predicted labels (i.e., 0 or 1 converted after softmax activation in Fig. 4b) and ground-truth labels, can be calculated after all weights are updated in each epoch. In summary, weights in matrix Y for any kernel can be updated iteratively according to above equations and stop updating until the loss between two epochs changes in a meagre range.

10. Computational prediction of small molecule inhibitors

In this section, we introduce the computational techniques for discovering small molecules inhibitors of miRNA and protein targets, which includes RNA motif-searching prediction, association prediction of small molecules and miRNAs (SM-miRs), regulation type prediction of associated SM-miRs, and drug target interaction (DTI) prediction. Most of the prediction processes require the expansion of data as input by indicating unknown SM-miR or drug-target relationships and deep learning modelling needs a feature engineering process, which are shown in sections 10.1.1 and 10.1.2, respectively.

10.1. Data augmentation and feature preparation

10.1.1. The guilt-by-association rule

Biological systems are often described as complex networks [142, 143], in which biological entities such as small molecules, proteins, and genes are represented as nodes and the relationships between them are represented as edges. These biological networks can be further categorized into homogeneous and heterogeneous networks [144]. In detail, biological entities of one kind (e.g., drugs) are connected with one another into the homogeneous network, such as the drug-drug interaction network, while biological entities of different kinds (e.g., drugs and targets) are inter-connected into the heterogeneous network in which its two representations, bipartite [145] (e.g., the drug-target interaction network) and tripartite [146] (e.g., the drug-target-disease association network or the drug-target-response network) graphs, are very commonly seen in drug discovery fields [147–149]. While heterogeneous networks usually contain a hefty volume of nodes, the relationships between these heterogeneous nodes can largely be unknown [146, 150]. One approach for deducing such unknown relationships is the guilt-by-association (GBA) approach. It operates under the assumption that if biological entities with unknown function have an interaction profile similar to those of known function, they may function similarly [151–153]. This implies that if drug A is known to treat a disease, the GBA-deduced drugs similar to drug A may also potentially be used to treat this disease. This approach has been demonstrated very useful in drug discovery [146,150,154,155]. Often, the relationships in heterogeneous networks are deduced by utilizing information from a medley of

homogeneous networks [144]. For example, if drug A is linked to protein X in a drug-protein interaction network, a new link between protein X and drug B in this network is required to be built by comparing the interaction profile between drug A and drug B in an external drug-drug interaction network where the two drugs coexist. Similarly, finding new protein targets of a drug in this drug-protein interaction network requires the information from an external protein-protein interaction network. The drug-drug interaction network or the protein-protein interaction network can be established by using those metrics that reflect its sequence, structural, functional properties, such as Tanimoto coefficients (for evaluating the chemical similarity between drugs) [156], sequence identity (for evaluating the sequence similarity between proteins) [27], or the direct experimental evidence to support interactions between drugs or proteins [157].

10.1.2. Feature engineering for ribonucleic acids, amino acids, and compounds

Computational studies of SM-miRNA interactions and DTIs often require a great amount of feature engineering work to characterize the biophysical and biochemical properties of miRNAs, proteins, and compounds. Compositions of miRNAs and proteins are well-suited for machine learning approaches taking the entire molecular sequences as input, since this kind of feature can yield a fixed-length feature encoding vector regardless of the issue on the input sequences of varying lengths. The compositional features are, in short, the frequencies of k -mers, where k usually takes 1, 2, or 3 at better, leading to 20, 400, 8000 dimensions for proteins for example. Tools that are powerful for generating these features include propy [158], PyDPI [159], PyBioMed [160], and PyFeat [161] available on Python interfaces, and iFeature [162], iLearn [163], iLearnPlus [164], and iFeatureOmega [165] available on webserver interfaces. In contrast, compounds have more complex features suitable for machine learning modelling, for example, fingerprints for compound's structural representations or partition coefficient (log P) and molar refractivity [166] for compound's physicochemical properties. There is, perhaps, no caveat to utilize diverse compound features for interaction prediction problems. The RDKit tool is widely applied in this aspect [167].

10.2. Computational design of lead compounds based on miRNA motifs

Inforna is a web server that provides computational design of lead small molecule compounds that target RNA molecules. It works mainly as an algorithm-embedded, expert-curated database of RNA motif-small molecule relationships [21]. With only an input RNA sequence of any length, Inforna undergoes the two stages of prediction-based comparisons: predicting its secondary structures and predicting the fitness of an RNA motif-small molecule interaction through the StARTS scoring scheme to produce lead compounds [22]. The concept is that if the predicted secondary structures of the input RNA share a high similarity with the RNA motif paired with a small molecule, the input RNA can be linked to that small molecule [168]. These deposited RNA motif-small molecule relationships are rendered reliable since they are derived from high-quality microarray-based screening experiments [169]. Several experiments provide proof of the accuracy of computation-led selection of lead compounds [170,171]. For example, Suresh et al. reported that a compound prioritized by Inforna to target the Dicer processing site in pre-miR-21 resulted in a moderate inhibitory effect in treating triple-negative breast cancer [172]. Later, Liu et al. showed in another experiment that a Inforna-designed compound successfully targeted a motif in the Dicer processing sites of several pri-miR-17-92 family members (miR-17, miR-18a, and miR-20a), with the potential to treat polycystic kidney disease [173].

10.3. Prediction of associations between small molecules and miRNAs

A number of models have been proposed to predict the SM-miR

associations based on network/matrix-based inference approaches and machine learning techniques. As of now, we have collected from the literature 20 methods, of which 12 are categorized into the first category, including BNNRSMMA [174], EKRRSMMA [175], WKNKN [32], RWNS [176], GISMMA [177], SMANMF [178], SLHGISMMA [179], TLHNSMMA [180], HSSMMA [181], CLDISMMA [182], SMMART [183], RWR [184], and SNMFSMMA [185], and 7 belong to the latter group, including ELDMA [186], PSRR [187], DAESTB [188], RFSMMA [189], SMAJL [190], Abdelbaky et al. work [191], and Jamal et al. work [192]. The categorization is not restricted because some of the network/matrix-based methods, e.g., the SMMART work, may also involve an optimization process yet without a negative training dataset, thereby being excluded from the machine learning category. Some of the methods, e.g., the Abdelbaky et al. work, offer disease inference for outreach to showcase the application of drug discovery for a particular disease. The golden standard database that is most prevalently used among these methods is SM2miR [193], released one decade ago. In an effort to make a fair performance comparison, researchers have created two publicly available datasets of well-curated SM-miR associations, which most studies named Dataset 1 and Dataset 2. But the SM-miR association database is not only confined to SM2miR. For example, Jamal et al. initiated a collection of more than 300 thousand small-molecule modulators and/or inhibitors of miR-21 via the PubChem AID 2289 by using the quantitative HTS (qHTS) assay [192]. Through an activity scoring scheme, these small molecules are tagged as active or inactive, which are then subject to data mugging and feature engineering and finally delivered to the Naïve Bayes [194] and random forest [195] learning algorithms for inferring the inhibition likelihoods of miRNA targets. The SM-miR association prediction field is undergoing rapid progress especially in the past few years, in terms of the surge in the number of method-based publications and the predictive ability. According to the performance comparison in these studies, the two newest methods, WKNKN and DAESTB, improved AUC values from 0.863 to 0.986 on the two datasets. As can be seen above, this field lacks machine learning implementations, especially deep learning implementations. The primary reason is that in order to avoid overfitting brought about by the label imbalance problem [196,197], machine learning algorithms usually seek to be trained on label ratio-balanced samples, but the current SM-miR databases are filled with known association samples instead of non-association samples. What is more important, manually/computationally added putative negative samples (i.e., non-associations) may harm the performance of machine learning models, as discussed in Ref. [183].

In addition to the direct connections between drugs and miRNAs, some studies have incorporated functional data on drugs, such as drug responses, to model their relationship with miRNAs. For instance, Huang et al. developed the GCMDR approach to predict the associations between drug resistance and miRNAs [198], while Li et al. employed a heterogeneous network framework to predict the relationships between anticancer drug responses and miRNAs [199]. Unlike the SM-to-miRNA regulation direction, which involves a direct physical interaction between them, both studies explored how drug responses are modulated by miRNAs. This line of research is referred to as miRNA pharmacogenomics, which aims to understand the mechanism of drug action on diseases by linking drugs, miRNAs, and genes together [200]. Specifically, the rationale is that the response of a drug is affected by miRNAs through the dysregulation of the expression of genes whose protein products bind to this drug. These kinds of triplets (miRNAs- > genes- > drugs), where the drug response lies at the very last link of the chain reaction, indirectly link miRNAs and drugs, as indicated by Rukov et al. [201].

10.4. Prediction of regulation type of associated small molecules and miRNAs

Computational approaches have advanced the discovery of drugs

targeting miRNAs for disease treatment, following the progress made in predicting SM-miR associations [202]. Summarized in this review [203], drugs can be repositioned efficiently by an *in-silico* gauge of the presence or absence of SM-miR associations. The expression of miRNAs can be altered either upwards or downwards by binding with different small molecules, but currently, this information seems not to be seamlessly integrated with the later computational drug discovery process that combines with the disease knowledge. For example, the pathogenesis of a disease is related to the upregulation of the expression of a miRNA. Then, we pursue whether there are small molecules that are able to downregulate the miRNA expression. Knowing the importance of the SM-miR regulation types to the development of potential small molecule therapeutics [54,61,204,205], we have developed two computational tools, DeepsmirUD and DeeplncUD (manuscript in preparation), to predict the regulation types (cf. we referred to as regulatory effects in the two studies) of small molecules for altering the expression of miRNAs and lncRNAs, respectively. It is noted that a SM-miR pair taken as input to DeepsmirUD has to be an experimentally-verified or predicted association. After then, with the regulation type-specific information, miRNA expression signatures induced by small molecules were used to couple with disease-related miRNAs to find potential drugs by computing connectivity scores (details summarized early in this review). The data availability is of concern to the computational method development. We downloaded two data resources, SM2miR [193] and D-lnc [206], which made it possible to access the machine learning approaches in miRNAs and lncRNAs, respectively. The regulation type prediction has not reached other types of ncRNAs as we did not collect a hoard of interaction data that suffices to train a machine learning model.

10.5. Prediction of interactions between drugs and protein targets (DTIs)

10.5.1. Databases of DTIs

As high-throughput experimental techniques are being increasingly applied for screening drug candidates that activate or inhibit protein targets related to a certain disease(s), there have been a growing number of well-established databases of DTIs with a concomitant increase in the number of DTIs [207], including DrugBank [84], TTD [208], SuperTarget [209]. The information about other available webservers managing the DTI data has been well-documented in these three reviews [33, 210,211]. We mainly focus on the analysis of the DrugBank database due to its pervasive applications in computational DTI studies. The data is compiled in DrugBank at different levels of granularity, such as SNP-associated drug effects, drug-drug interactions (DDIs), DTIs, etc. Following a major overhaul in its 5th version, DrugBank has substantially expanded its data content, with several categories experiencing at least a 1-fold increase (e.g., 600% for DDIs) and a few recent additions, such as drug clinical trial data. While the precise number of DTIs is not explicitly revealed in this version change, statistics calculated based on the two most recent versions (5.1.8 and 5.1.9 downloaded through the target sequence mark) indicate a slight increase in the number of DTIs, with a 3.13% increase for FDA-approved-drugs and a relatively unchanged ratio for experimental-drugs (i.e., preclinical or animal testing phase) (Table 4). Furthermore, there has been a noteworthy increase of approximately 6.46% in the number of protein targets that have been identified as being targeted by FDA-approved drugs. For example, there are 57 new drugs tagged as FDA-approved. After removing repeated records and those target sequences containing non-standard amino acid symbols, the number of DTIs decreases 0.1-fold for FDA-approved drugs and 0.05-fold for experimental drugs. Overall, the recent DrugBank database maintains ~10,000 DTIs per each category, which serve as adequate materials for computational modelling and analysis. Based on the DrugBank database, we generated a DTI dataset to train a predictor, Drutai, in order to determine how accurately deep neural networks can predict DTIs using molecular sequences alone (see Supplementary Materials). Then, we use Drutai to predict DTIs when targets and drugs are

Table 4
Number of drugs, targets, and DTIs in DrugBank.

Version	Condition	Approved			Experimental		
		Target	Drug	Interactions	Target	Drug	Interactions
5.1.8	raw	2694	2170	11058	2984	4549	8698
	non-repeated non-UNK	2682	2167	10738	2970	4548	8549
	PL < 3000	2484	1839	9776	2851	4481	8306
5.1.9	raw	2868	2227	11415	2983	4551	8692
	non-repeated non-UNK	2856	2224	11083	2969	4550	8543
	PL < 3000	2650	1872	10,087	2837	4482	8280

Note. UNK: unknown characters out of 20 standard amino acid symbols. PL: protein length.

novel (Figs. 5 and 6, see <https://aidrugud.github.io/drutai> for interactive graphs), respectively.

10.5.2. Computational tools for predicting DTIs

The development of therapeutics often involves using druglike small molecules to bind to protein targets for inhibition or activation [212]. Consequently, predicting DTIs has garnered significant attention as it can greatly reduce the experimental efforts required to test the therapeutic potential of small molecules [23]. Since then, improving the prediction of interactions between drugs and protein targets is one of the most promising fields in computational drug discovery [213]. With accumulated computational efforts, this research area has now been overwhelmed by numerous predictors, which are mainly categorized into sequence-based and structure-based predictors. There are too many sequence-based approaches to enumerate, such as Yu et al. work [214], EnsemDT [215], and DTINet [216]. Recent years have also seen a rising number of deep learning methods, many of which conceive novel schemes of taking features of DTIs as input to deep neural network architectures [217,218]. For example, the learning scheme implemented by DeepDTA [219] is that a small molecule and a protein are fed into two separate CNNs to be trained in parallel for learning their latent representations, which are then concatenated into one feature vector that is finally translated by conventionally a fully-connected neural network into the propensity of it being in interaction or non-interaction. Likewise, other methods that adopt a similar idea include DeepAffinity [220], DeepPurpose [221], MDDeePred [222], MINN-DTI [223], Multi-DTI [217], and HyperAttentionDTI [224]. It is worth mentioning that in the MDDeePred, an image-like object that hierarchically represents protein features through channels is used as input to a deep learning architecture, while in the MINN-DTI work, the distance map is leveraged to improve the representation of target proteins. Besides, the utility of graphs involved in the sequence-based prediction has been demonstrated by a few studies [225–227]. The crux of the plan is to represent a chemical structure as a graph or add relevant protein-protein or drug-drug interaction networks, which fit into graph convolutional neural networks (for review, see Ref. [228]). Comparatively, the structure-based prediction has yet to be popularized mainly due to the availability of protein structures, the additional computation on structures, and the lack of the effective plans of the incorporation of protein structures and learning algorithms. Existing studies have given several examples to incorporate protein structures into the prediction DTIs, e.g., protein sequence annotations in domain resolution [229], graph representations of binding pockets [230], and predicted residue contact maps [231]. The DrugBank database [84], as an internationally trusted source of DTIs, is involved in a large part in the development of many methods, such as DrugE-Rank [232], LASSO-DNN [156], DeepDTIs [233], DrugR+ [234]. To illustrate the applicability of the prediction tools, DTIs formed with novel drugs or novel targets are curated as in Refs. [235–237].

10.6. Network analysis of the connectivity scoring scheme using drug-mediated miRNA profiles predicted by DeepsmirUD

To illustrate the utility of our piloted scheme based on connectivity scores to reposition conventional drugs or discover new drugs, we carried out a drug-miRNA-disease network analysis. To this end, we extracted from DeepsmirUD-Web [27] the drug-mediated miRNA profiles with absolute values of predicted regulatory effects (i.e., likelihoods of upregulation or downregulation types) of greater than 0.95 and the drug-disease associations with absolute values of connectivity scores of greater than 0.5.

We present five case studies shown in Fig. 7. In ellipses 1–3, our network analysis successfully predicts the capability of gemcitabine [238], lenalidomide [239], and paclitaxel [240] to treat breast carcinoma, glioblastoma, and pancreatic cancer, respectively, as highlighted by the negative weights ranging between -1 and 0 . Considering its several side effects, such as nausea, fatigue, low blood cell counts, and allergic reactions [241], gemcitabine should cautiously be used for treating breast carcinoma. However, the network created using the connectivity scoring scheme is also full of wrong indications or links that remain unknown. Apart from a successfully predicted true link between piperine and osteosarcoma [242], ellipse 4 gives a false or currently unknown link between promethazine and osteosarcoma. Moreover, 5-fluorouracil, as a chemotherapy medication, has a potential therapeutic effect on colorectal cancer (CRC) [243]. However, there is the other way around in ellipse 5, indicated by the positively connected strength of 5-fluorouracil with CRC. It is noted that due to its chemoresistance, 5-fluorouracil is suggested to be better used in conjunction with other natural or synthetic compounds for the co-treatment of CRC.

11. Discussion

This review provides an overview of miRNA biogenesis, functions, and disease implications, from both biological and computational perspectives. Additionally, it presents the current advancements in computational strategies for the discovery of small molecule drugs of miRNA targets. By focusing on the small molecule-mediated targeting of miRNAs, this review is structured in a concise format to summarize experimental technologies for drug screening in tandem with computational strategies. To link drugs and diseases, we begin by stressing the importance of associations of miRNAs with them, respectively, since aberrant biomolecules (say, genes, transcripts, and proteins) constitute the genetic and phenotypic basis that underlies the pathogenesis of diseases [14], and these molecules provide the structural basis that accounts for exerting cellular functions in somewhat of a high ratio by interacting with small molecule compounds [18]. In particular, compare to targeting proteins, targeting miRNAs with druglike small molecules is detailed by a large margin of this review. This is because the development of miRNA-targeted small molecule inhibitors has burgeoned a new field, where drugs are found through the post-transcriptional regulation mechanism of miRNAs rather than using protein information. This way is more straightforward owing to the miRNAs' direct regulation control of deregulated gene expression that triggers diseases and, furthermore,

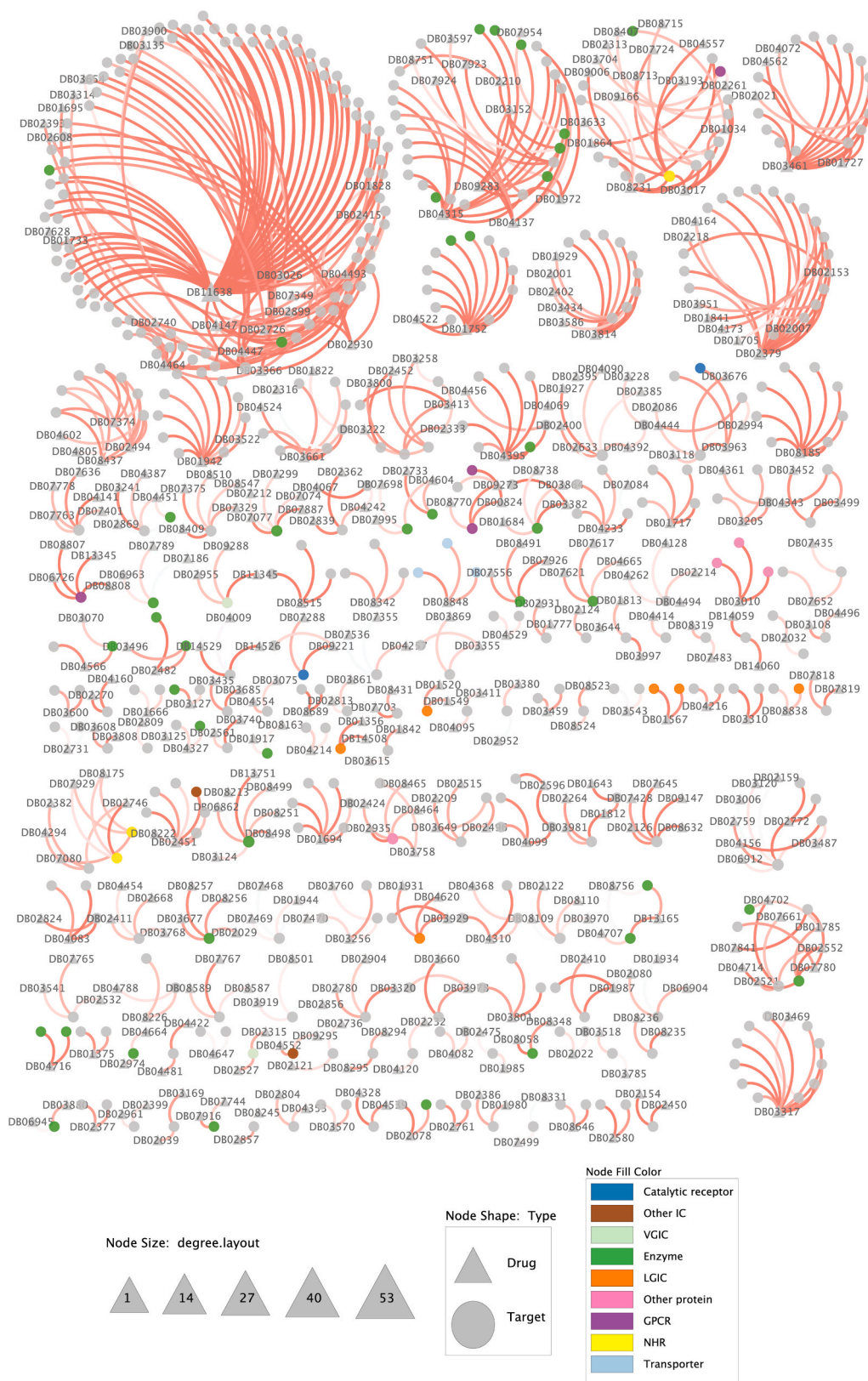
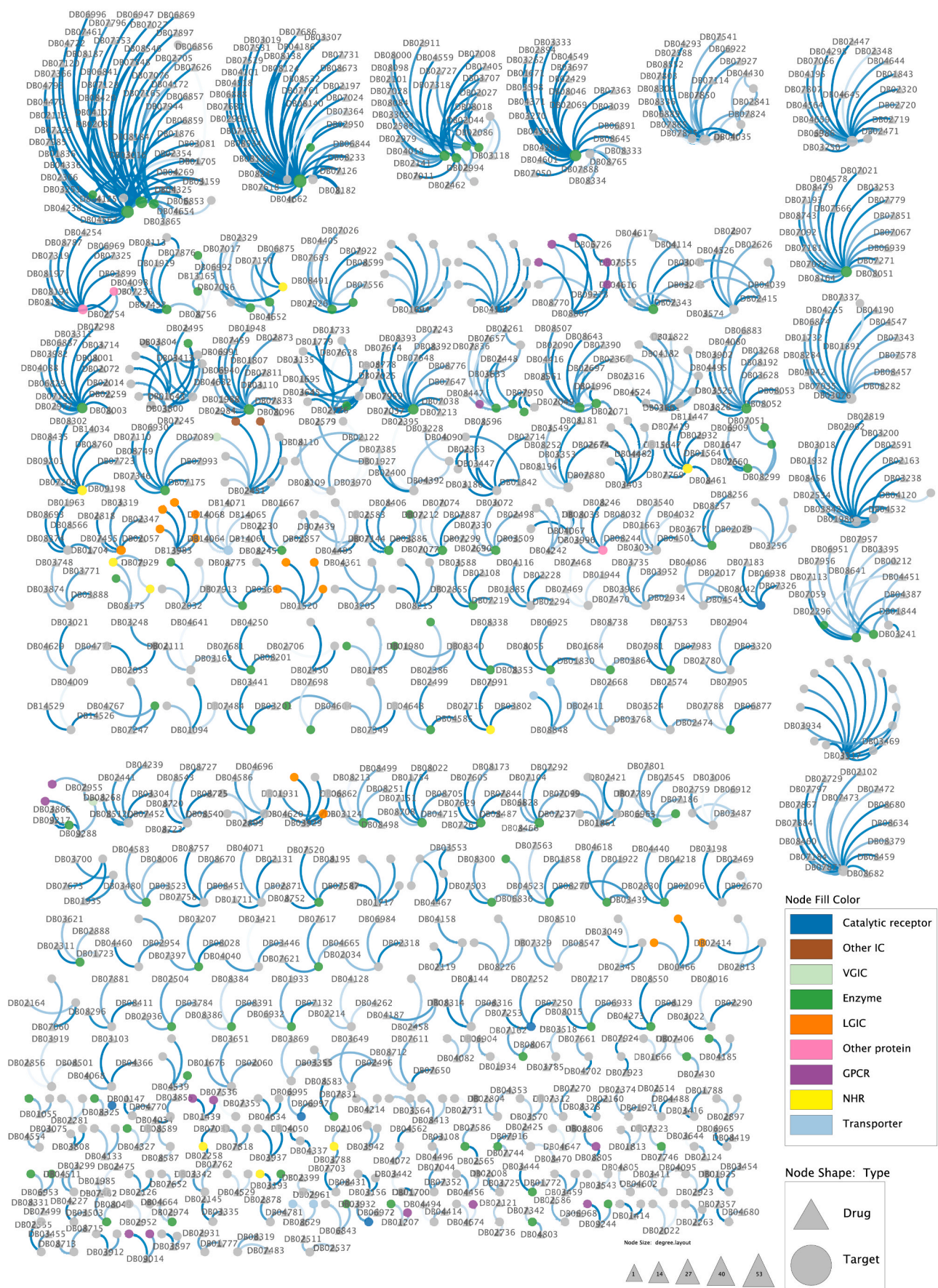


Fig. 5. Network of DTIs formed with novel therapeutic targets. Each line represents ground-truth evidence of an interaction between a protein target and a drug. The connection strength of each interaction is indicated by using the Drutai-predicted DTI probability. Drugs and protein targets are represented by the triangle and the circle, respectively. The size of the triangle of a drug becomes bigger if the drug is identified to connect with more targets (see <https://aidrugud.github.io/drutai> for interactive graphs). Human transmembrane proteins that are of pharmacological significance are annotated based on the GtoPdb database [260] using 9 functional groups: *Transporter*, *GPCR* (G-protein-coupled receptor), *Enzyme*, *Catalytic receptor*, *LGIC* (ligand-gated ion channel), *VGIC* (voltage-gated ion channel), *NHR* (nuclear hormone receptors), *Other IC* (other ion channel), and *Other protein*.

the control of protein noises by targeting lowly expressed genes, as evidenced in an experiment on mouse embryonic stem cells [244]. Specifically, we elaborate the crucial role of miRNAs in developing cancer therapeutics from their biogenesis, functions, and experimental determination, eventually to the clinical applications, in that their

aberrant expression is thought to affect cell proliferation, differentiation, and apoptosis. miRNA-targeted small molecule inhibitors have been proven to be useful to antagonize oncogenic pathways. Creating a solution for drug discovery is inextricably linked with experiment-led and computation-assisted efforts in concert. Experimentally, the



(caption on next page)

Fig. 6. Network of DTIs formed with novel drugs. Each line represents ground-truth evidence of an interaction between a protein target and a drug. The connection strength of each interaction is indicated by using the Drutai-predicted DTI probability. Drugs and protein targets are represented by the triangle and the circle, respectively. The size of the triangle of a drug becomes bigger if the drug is identified to connect with more targets (see <https://aidrugud.github.io/drutai> for interactive graphs). Human transmembrane proteins that are of pharmacological significance are annotated based on the GtoPdb database [260] using 9 functional groups: *Transporter*, *GPCR* (G-protein-coupled receptor), *Enzyme*, *Catalytic receptor*, *LGIC* (ligand-gated ion channel), *VGIC* (voltage-gated ion channel), *NHR* (nuclear hormone receptors), *Other IC* (other ion channel), and *Other protein*.

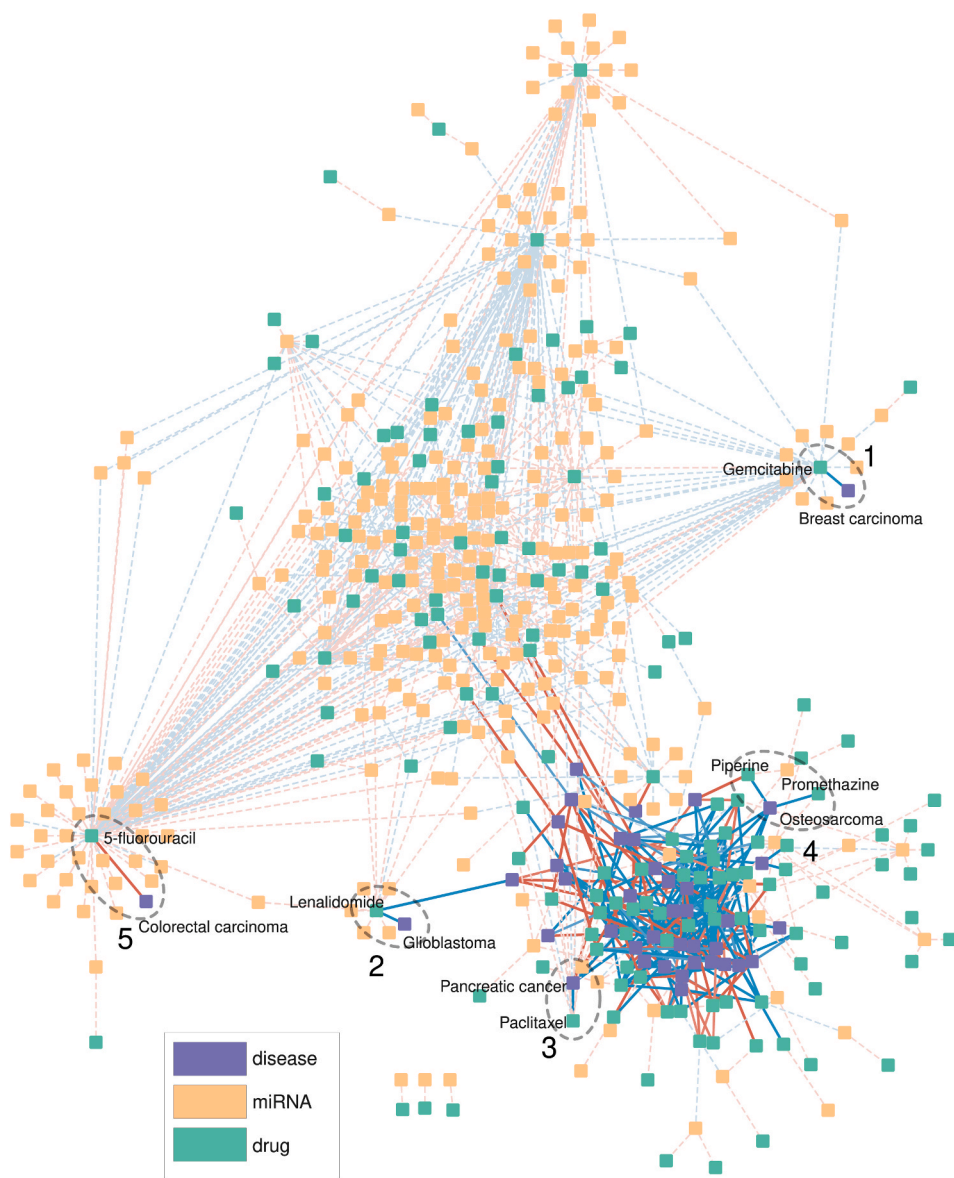


Fig. 7. Network analysis based on small molecule-miRNA, miRNA-disease, and small molecule-disease relationships extracted from DeepSmirUD-Web. The five ellipses highlight the predicted druglike potential of five small molecules (gemcitabine, lenalidomide, paclitaxel, piperine, and 5-fluorouracil) for treating breast carcinoma, glioblastoma, pancreatic cancer, osteosarcoma, and colorectal cancer, respectively. The blue line indicates the druglike potential of a small molecule while the red line indicates the ability of a small molecule to be supportive of a disease. The blue and red dashed lines indicate the downregulation and upregulation type of small molecules on modulating miRNA expression, respectively. Small molecule-disease relationships are linked based on connectivity scores [28].

discovery process with current technologies is accompanied by a time-consuming preparation of a bespoke chemical library for high-throughput screening, which makes drug determination challenging. Rather than substituting the existing models, computational techniques should play a supporting yet sometimes dominant role in accelerating the duration of drug discovery. Overall, the clinical development of miRNA drugs is still considerably enroute.

The state-of-the-art computational capabilities have actively been used in the field of drug discovery and, in fact, can be layered into different stages during this course, including drug design, drug indication, drug repositioning, drug interactions, etc. For example, in early drug discovery, they can participate in the prediction of drug interactions or affinities with protein or miRNA targets for screening lead

compounds, which can partly reduce the heavy workloads throughout HTS. Methodologically, these computational techniques can roughly be divided into two groups in light of whether the algorithms, themselves, demand a certain optimization process for generating intelligent models. For example, the Drutai tool demonstrated alongside this review was assembled as a result of a series of optimization iterations. Importantly, similarity-based inference techniques, many of which do not necessitate an optimization procedure for parameter estimation, are prevalently used in this field. As in Ref. [183], for example, the drug similarity is calculated as the weighted arithmetic mean of the clinical and structural similarities and the miRNA similarity is estimated based on the gene functional similarity and the cosine similarity. Yet another example is shown in Ref. [22] that the drug's Tanimoto similarity is employed in

search of lead compounds. We found that the sequence-based prediction of DTIs has been in the groove by achieving 85–95% AUC values, implying that the current volume of DTIs in the DrugBank database suffices to generate deep learning models to be used in practice. In addition, it is very interesting to watch further cases or new modalities for validating the efficacy of the miRNA drugs discovered based on DeepSmirUD-predicted regulation types, disease-miRNA relationships, and connectivity scores.

Established studies have extensively benchmarked and discussed the biogenesis [245], structures [15], functions [42], targeting properties [4], silencing [246] of miRNAs, which, together, give insight into the pathogenesis of miRNA-regulated pathways [14]. These biological findings have yet to be tightly integrated with computational efforts. Starkly, the field of miRNAs is greatly devoid of functional annotations in their structures [247], due to a much more pronounced paucity of 3D native structures than those of proteins. The bulk of the ongoing efforts is being given to the design of covariation-based tools [248] for prediction [15]. From a limited number of case reports, we gained an understanding about the mechanism of the miRNA targeting that is specifically and selectively carried out in structured regions, like in the stem-loop hairpin secondary structure. Either or both the prediction-assisted and the experiment-led annotations of function sites are assumed to play a nonnegligible role in enabling the downstream analysis for computation-assisted structure-based drug design and discovery, provided that these computationally elaborated structures can be put on a par with those experimentally resolved ones. In this aspect, with many experimentally resolved miRNA structures becoming available, the information on specific functional sites (binding/interaction/inhibition) derived from these structures will be accumulated and can be used to train cutting-edge deep learning models, which will supply accurate prediction-assisted annotations of function sites in miRNAs on a large scale.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Drutai is available at <https://github.com/2003100127/drutai>, which is used for predicting the interactions between drugs and protein targets

Acknowledgement

This study is supported by the Medical Research Council (MRC) career development fellowship (MR/V010182/1) to Adam P. Cribbs. Xia Wang is supported by the Chinese Universities Scientific Fund (2452022255) and the Natural Science Foundation of Shaanxi Province (2023-JC-YB-177).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmech.2023.115500>.

References

- [1] H. Beck, M. Härter, B. Haß, C. Schmeck, L. Baerfacker, Small molecules and their impact in drug discovery: a perspective on the occasion of the 125th anniversary of the Bayer Chemical Research Laboratory, *Drug Discov. Today* 27 (2022) 1560–1574, <https://doi.org/10.1016/j.drudis.2022.02.015>.
- [2] R. Offringa, L. Kötznér, B. Huck, K. Urbahns, The expanding role for small molecules in immuno-oncology, *Nat. Rev. Drug Discov.* 21 (2022) 821–840, <https://doi.org/10.1038/s41573-022-00538-9>.
- [3] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.* 46 (2017) D1074–D1082, <https://doi.org/10.1093/nar/gkx1037>.
- [4] Z. Li, T.M. Rana, Therapeutic targeting of microRNAs: current status and future challenges, *Nat. Rev. Drug Discov.* 13 (2014) 622–638, <https://doi.org/10.1038/nrd4359>.
- [5] A. Donlic, A.E. Hargrove, Targeting RNA in mammalian systems with small molecules, *WIREs RNA* 9 (2018) e1477, <https://doi.org/10.1002/wrna.1477>.
- [6] N.F. Rizvi, John P. Santa Maria Jr., A. Nahvi, J. Klappenbach, D.J. Klein, P. J. Curran, M.P. Richards, C. Chamberlin, P. Saradjian, J. Burchard, R. Aguilar, J. T. Lee, P.J. Dandliker, G.F. Smith, P. Kutchukian, E.B. Nickbarg, Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space, *SLAS Discovery* 25 (2020) 384–396, <https://doi.org/10.1177/2472555219885373>.
- [7] C.-K. Huang, S. Kafert-Kasting, T. Thum, Preclinical and clinical development of noncoding RNA therapeutics for cardiovascular disease, *Circ. Res.* 126 (2020) 663–678, <https://doi.org/10.1161/CIRCRESAHA.119.315856>.
- [8] K.D. Warner, C.E. Hajdin, K.M. Weeks, Principles for targeting RNA with drug-like small molecules, *Nat. Rev. Drug Discov.* 17 (2018) 547–558, <https://doi.org/10.1038/nrd.2018.93>.
- [9] T.L. Nero, C.J. Morton, J.K. Holien, J. Wielens, M.W. Parker, Oncogenic protein interfaces: small molecules, big challenges, *Nat. Rev. Cancer* 14 (2014) 248–262, <https://doi.org/10.1038/nrc3690>.
- [10] E. Anastasiadou, L.S. Jacob, F.J. Slack, Non-coding RNA networks in cancer, *Nat. Rev. Cancer* 18 (2018) 5–18, <https://doi.org/10.1038/nrc.2017.99>.
- [11] X.-D. Fu, Non-coding RNA: a new frontier in regulatory biology, *Natl. Sci. Rev.* 1 (2014) 190–204, <https://doi.org/10.1093/nsr/nwu008>.
- [12] D. Rearick, A. Prakash, A. McSweeney, S.S. Shepard, L. Fedorova, A. Fedorov, Critical association of ncRNA with introns, *Nucleic Acids Res.* 39 (2010) 2357–2366, <https://doi.org/10.1093/nar/gkq1080>.
- [13] M. Matsui, D.R. Corey, Non-coding RNAs as drug targets, *Nat. Rev. Drug Discov.* 16 (2017) 167–179, <https://doi.org/10.1038/nrd.2016.117>.
- [14] M. Esteller, Non-coding RNAs in human disease, *Nat. Rev. Genet.* 12 (2011) 861–874, <https://doi.org/10.1038/nrg3074>.
- [15] J.L. Childs-Disney, X. Yang, Q.M.R. Gibaut, Y. Tong, R.T. Batey, M.D. Disney, Targeting RNA structures with small molecules, *Nat. Rev. Drug Discov.* 21 (2022) 736–762, <https://doi.org/10.1038/s41573-022-00521-4>.
- [16] P. Wu, Inhibition of RNA-binding proteins with small molecules, *Nat. Rev. Chem.* 4 (2020) 441–458, <https://doi.org/10.1038/s41570-020-0201-4>.
- [17] M. Winkle, S.M. El-Daly, M. Fabbri, G.A. Calin, Noncoding RNA therapeutics — challenges and potential solutions, *Nat. Rev. Drug Discov.* 20 (2021) 629–651, <https://doi.org/10.1038/s41573-021-00219-z>.
- [18] M. Cully, Small molecules targeting a tertiary RNA structure fight fungi, *Nat. Rev. Drug Discov.* 17 (2018) 864, <https://doi.org/10.1038/nrd.2018.205>.
- [19] P. Csermely, T. Korcsmáros, H.J.M. Kiss, G. London, R. Nussinov, Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review, *Pharmacol. Ther.* 138 (2013) 333–408, <https://doi.org/10.1016/j.pharmthera.2013.01.016>.
- [20] M.D. Disney, Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine, *J. Am. Chem. Soc.* 141 (2019) 6776–6790, <https://doi.org/10.1021/jacs.8b13419>.
- [21] S.P. Velagapudi, S.M. Gallo, M.D. Disney, Sequence-based design of bioactive small molecules that target precursor microRNAs, *Nat. Chem. Biol.* 10 (2014) 291–297, <https://doi.org/10.1038/nchembio.1452>.
- [22] M.D. Disney, A.M. Winkelsas, S.P. Velagapudi, M. Southern, M. Fallahi, J. L. Childs-Disney, Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs, *ACS Chem. Biol.* 11 (2016) 1720–1728, <https://doi.org/10.1021/acschembio.6b00001>.
- [23] H. Ding, I. Takigawa, H. Mamitsuka, S. Zhu, Similarity-based machine learning methods for predicting drug–target interactions: a brief review, *Briefings Bioinf.* 15 (2013) 734–747, <https://doi.org/10.1093/bib/bbt056>.
- [24] T. Gaudet, B. Day, A.R. Jamasb, J. Soman, C. Regep, G. Liu, J.B.R. Hayter, R. Vickers, C. Roberts, J. Tang, D. Roblin, T.L. Blundell, M.M. Bronstein, J. P. Taylor-King, Utilizing graph machine learning within drug discovery and development, *Briefings Bioinf.* 22 (2021), <https://doi.org/10.1093/bib/bbab159>.
- [25] Z. Wu, W. Li, G. Liu, Y. Tang, Network-based methods for prediction of drug–target interactions, *Front. Pharmacol.* 9 (2018), <https://doi.org/10.3389/fphar.2018.01134>.
- [26] Y. Zhang, T. Ye, H. Xi, M. Juhas, J. Li, Deep learning driven drug discovery: tackling severe acute respiratory syndrome coronavirus 2, *Front. Microbiol.* 12 (2021), <https://doi.org/10.3389/fmicb.2021.739684>.
- [27] J. Sun, J. Ru, L. Ramos-Mucci, F. Qi, Z. Chen, S. Chen, A.P. Cribbs, L. Deng, X. Wang, DeepSmirUD: prediction of regulatory effects on microRNA expression mediated by small molecules using deep learning, *Int. J. Mol. Sci.* 24 (2023), <https://doi.org/10.3390/ijms24031878>.
- [28] K. Samart, P. Tuyishime, A. Krishnan, J. Ravi, Reconciling multiple connectivity scores for drug repurposing, *Briefings Bioinf.* 22 (2021), <https://doi.org/10.1093/bib/bbab161>.
- [29] J. Lamb, E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S.J. Haggarty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, T. R. Golub, The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease, *Science* 313 (1979) 1929–1935, <https://doi.org/10.1126/science.1132939>, 2006.

- [30] A. Musa, L.S. Ghorai, S.-D. Zhang, G. Glazko, O. Yli-Harja, M. Dehmer, B. Haibe-Kains, F. Emmert-Streib, A review of connectivity map and computational approaches in pharmacogenomics, *Briefings Bioinf.* 19 (2017) 506–523, <https://doi.org/10.1093/bib/bbw112>.
- [31] N. El-Hachem, D.M.A. Gendoo, L.S. Ghorai, Z. Safikhani, P. Smirnov, C. Chung, K. Deng, A. Fang, E. Birkwood, C. Ho, R. Isserlin, G.D. Bader, A. Goldenberg, B. Haibe-Kains, Integrative cancer pharmacogenomics to infer large-scale drug taxonomy, *Cancer Res.* 77 (2017) 3057–3069, <https://doi.org/10.1158/0008-5472.CAN-17-0096>.
- [32] S.-H. Wang, C.-C. Wang, L. Huang, L.-Y. Miao, X. Chen, Dual-Network Collaborative Matrix Factorization for predicting small molecule-miRNA associations, *Briefings Bioinf.* 23 (2021), <https://doi.org/10.1093/bib/bbab500>.
- [33] M. Bagherian, E. Sabeti, K. Wang, M.A. Sartor, Z. Nikolovska-Coleska, K. Najarian, Machine learning approaches and databases for prediction of drug–target interaction: a survey paper, *Briefings Bioinf.* 22 (2020) 247–269, <https://doi.org/10.1093/bib/bbz157>.
- [34] F. Boniolo, E. Dorigatti, A.J. Ohnmacht, D. Saur, B. Schubert, M.P. Menden, Artificial intelligence in early drug discovery enabling precision medicine, *Expert Opin. Drug Discov.* 16 (2021) 991–1007, <https://doi.org/10.1080/17460441.2021.1918096>.
- [35] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, The rise of deep learning in drug discovery, *Drug Discov. Today* 23 (2018) 1241–1250, <https://doi.org/10.1016/j.drudis.2018.01.039>.
- [36] S. Lin, R.I. Gregory, MicroRNA biogenesis pathways in cancer, *Nat. Rev. Cancer* 15 (2015) 321–333, <https://doi.org/10.1038/nrc3932>.
- [37] E. Mogilyansky, I. Rigoutsos, The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease, *Cell Death Differ.* 20 (2013) 1603–1614, <https://doi.org/10.1038/cdd.2013.125>.
- [38] Y. Lu, J.M. Thomson, H.Y.F. Wong, S.M. Hammond, B.L.M. Hogan, Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells, *Dev. Biol.* 310 (2007) 442–453, <https://doi.org/10.1016/j.ydbio.2007.08.007>.
- [39] V. Patel, D. Williams, S. Hajarnis, R. Hunter, M. Pontoglio, S. Somlo, P. Igarashi, miR-17–92 miRNA cluster promotes kidney cyst growth in polycystic kidney disease, *Proc. Natl. Acad. Sci. USA* 110 (2013) 10765–10770, <https://doi.org/10.1073/pnas.1301693110>.
- [40] C. Chakraborty, A.R. Sharma, G. Sharma, S.-S. Lee, Therapeutic advances of miRNAs: a preclinical and clinical update, *J. Adv. Res.* 28 (2021) 127–138, <https://doi.org/10.1016/j.jare.2020.08.012>.
- [41] E.C. Lee, T. Valencia, C. Allerson, A. Schairer, A. Flaten, M. Yheskel, K. Kersjes, J. Li, S. Gatto, M. Takhar, S. Lockton, A. Pavlicek, M. Kim, T. Chu, R. Soriano, S. Davis, J.R. Androsavich, S. Sarwary, T. Owen, J. Kaplan, K. Liu, G. Jang, S. Neben, P. Bentley, T. Wright, V. Patel, Discovery and preclinical evaluation of anti-miR-17 oligonucleotide RGLS4326 for the treatment of polycystic kidney disease, *Nat. Commun.* 10 (2019) 4148, <https://doi.org/10.1038/s41467-019-11918-y>.
- [42] J.S. Mattick, P.P. Amaral, P. Carninci, S. Carpenter, H.Y. Chang, L.-L. Chen, R. Chen, C. Dean, M.E. Dinger, K.A. Fitzgerald, T.R. Gingeras, M. Guttman, T. Hirose, M. Huarte, R. Johnson, C. Kanduri, P. Kapranov, J.B. Lawrence, J. T. Lee, J.T. Mendell, T.R. Mercer, K.J. Moore, S. Nakagawa, J.L. Rinn, D. L. Spector, I. Ulitsky, Y. Wan, J.E. Wilusz, M. Wu, Long non-coding RNAs: definitions, functions, challenges and recommendations, *Nat. Rev. Mol. Cell Biol.* (2023), <https://doi.org/10.1038/s41580-022-00566-8>.
- [43] Y. Chen, Z. Li, X. Chen, S. Zhang, Long non-coding RNAs: from disease code to drug role, *Acta Pharm. Sin. B* 11 (2021) 340–354, <https://doi.org/10.1016/j.apsb.2020.10.001>.
- [44] P. Sparber, A. Filatova, M. Khantemirova, M. Skoblov, The role of long non-coding RNAs in the pathogenesis of hereditary diseases, *BMC Med. Genom.* 12 (2019) 42, <https://doi.org/10.1186/s12920-019-0487-6>.
- [45] Z. Bao, Z. Yang, Z. Huang, Y. Zhou, Q. Cui, D. Dong, LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases, *Nucleic Acids Res.* 47 (2018) D1034–D1037, <https://doi.org/10.1093/nar/gky905>.
- [46] N. Yanaihara, N. Caplen, E. Bowman, M. Seike, K. Kumamoto, M. Yi, R. M. Stephens, A. Okamoto, J. Yokota, T. Tanaka, G.A. Calin, C.-G. Liu, C.M. Croce, C.C. Harris, Unique microRNA molecular profiles in lung cancer diagnosis and prognosis, *Cancer Cell* 9 (2006) 189–198, <https://doi.org/10.1016/j.ccr.2006.01.025>.
- [47] S.E. Ahmadi, S. Rahimi, B. Zarandi, R. Chegeni, M. Safa, MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies, *J. Hematol. Oncol.* 14 (2021) 121, <https://doi.org/10.1186/s13045-021-01111-4>.
- [48] A. Esquela-Kerscher, F.J. Slack, Oncomirs — microRNAs with a role in cancer, *Nat. Rev. Cancer* 6 (2006) 259–269, <https://doi.org/10.1038/nrc1840>.
- [49] M.S. Ebert, P.A. Sharp, MicroRNA sponges: progress and possibilities, *RNA* 16 (2010) 2043–2050, <https://doi.org/10.1261/rna.2414110>.
- [50] T.C. Roberts, R. Langer, M.J.A. Wood, Advances in oligonucleotide drug delivery, *Nat. Rev. Drug Discov.* 19 (2020) 673–694, <https://doi.org/10.1038/s41573-020-0075-7>.
- [51] R. Garzon, G. Marcucci, C.M. Croce, Targeting microRNAs in cancer: rationale, strategies and challenges, *Nat. Rev. Drug Discov.* 9 (2010) 775–789, <https://doi.org/10.1038/nrd3179>.
- [52] M. Roos, U. Pradère, R.P. Ngondo, A. Behera, S. Allegrini, G. Civenni, J. A. Zagalak, J.-R. Marchand, M. Menzi, H. Towbin, J. Scheuermann, D. Neri, A. Caffisch, C. v. Catapano, C. Ciaudo, J. Hall, A small-molecule inhibitor of Lin28, *ACS Chem. Biol.* 11 (2016) 2773–2781, <https://doi.org/10.1021/acscchembio.6b00232>.
- [53] K. Cheng, X. Wang, H. Yin, Small-molecule inhibitors of the TLR3/dsRNA complex, *J. Am. Chem. Soc.* 133 (2011) 3764–3767, <https://doi.org/10.1021/ja111312h>.
- [54] S. Zhang, L. Chen, E.J. Jung, G.A. Calin, Targeting MicroRNAs with small molecules: from dream to reality, *Clin. Pharmacol. Ther.* 87 (2010) 754–758, <https://doi.org/10.1038/clpt.2010.46>.
- [55] M.G. Costales, H. Aikawa, Y. Li, J.L. Childs-Disney, D. Abegg, D.G. Hoch, S. P. Velagapudi, Y. Nakai, T. Khan, K.W. Wang, I. Yildirim, A. Adibekian, E. T. Wang, M.D. Disney, Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer, *Proc. Natl. Acad. Sci. USA* 117 (2020) 2406–2411, <https://doi.org/10.1073/pnas.1914286117>.
- [56] J. Manigrasso, M. Marcia, M. De Vivo, Computer-aided design of RNA-targeted small molecules: a growing need in drug discovery, *Chem* 7 (2021) 2965–2988, <https://doi.org/10.1016/j.chempr.2021.05.021>.
- [57] E.C. Morishita, Discovery of RNA-targeted small molecules through the merging of experimental and computational technologies, *Expert Opin. Drug Discov.* 0 (2022) 1–20, <https://doi.org/10.1080/17460441.2022.2134852>.
- [58] X. Fang, Y. Zheng, Y. Duan, Y. Liu, W. Zhong, Recent advances in design of fluorescence-based assays for high-throughput screening, *Anal. Chem.* 91 (2019) 482–504, <https://doi.org/10.1021/acs.analchem.8b05303>.
- [59] S.P. Velagapudi, Y. Luo, T. Tran, H.S. Haniff, Y. Nakai, M. Fallahi, G.J. Martinez, J.L. Childs-Disney, M.D. Disney, Defining RNA–small molecule affinity landscapes enables design of a small molecule inhibitor of an oncogenic noncoding RNA, *ACS Cent. Sci.* 3 (2017) 205–216, <https://doi.org/10.1021/acscentsci.7b00009>.
- [60] D.E. Scott, A.G. Coyne, S.A. Hudson, C. Abell, Fragment-based approaches in drug discovery and chemical biology, *Biochemistry* 51 (2012) 4990–5003, <https://doi.org/10.1021/bi3005126>.
- [61] R. Zhao, J. Fu, L. Zhu, Y. Chen, B. Liu, Designing strategies of small-molecule compounds for modulating non-coding RNAs in cancer therapy, *J. Hematol. Oncol.* 15 (2022) 14, <https://doi.org/10.1186/s13045-022-01230-6>.
- [62] C. Chakraborty, A.R. Sharma, G. Sharma, C.G.P. Doss, S.-S. Lee, Therapeutic miRNA and siRNA: moving from bench to clinic as next generation medicine, *Mol. Ther. Nucleic Acids* 8 (2017) 132–143, <https://doi.org/10.1016/j.omtn.2017.06.005>.
- [63] S. Zhang, Z. Cheng, Y. Wang, T. Han, The Risks of miRNA Therapeutics: in a Drug Target Perspective, *Drug Des. Devel. Ther.*, 2021, pp. 721–733.
- [64] S.P. Velagapudi, M.G. Costales, B.R. Vummidi, Y. Nakai, A.J. Angelbelo, T. Tran, H.S. Haniff, Y. Matsumoto, Z.F. Wang, A.K. Chatterjee, J.L. Childs-Disney, M. D. Disney, Approved anti-cancer drugs target oncogenic non-coding RNAs, *Cell Chem Biol* 25 (2018) 1086–1094.e7, <https://doi.org/10.1016/j.chembiol.2018.05.015>.
- [65] A. Akinc, M.A. Maier, M. Manoharan, K. Fitzgerald, M. Jayaraman, S. Barros, S. Ansell, X. Du, M.J. Hope, T.D. Madden, B.L. Mui, S.C. Semple, Y.K. Tam, M. Ciufolini, D. Witzigmann, J.A. Kulkarni, R. van der Meel, P.R. Cullis, The Onpatro story and the clinical translation of nanomedicines containing nucleic acid-based drugs, *Nat. Nanotechnol.* 14 (2019) 1084–1087, <https://doi.org/10.1038/s41565-019-0591-y>.
- [66] E. van Rooij, A.L. Purcell, A.A. Levin, Developing MicroRNA therapeutics, *Circ. Res.* 110 (2012) 496–507, <https://doi.org/10.1161/CIRCRESAHA.111.247916>.
- [67] L.A. Bravo-Vázquez, I. Medina-Ríos, L.D. Márquez-Gallardo, J. Reyes-Muñoz, F. I. Serrano-Cano, S. Pathak, A. Banerjee, A. Bandyopadhyay, A.K. Dutta, S. Paul, Functional implications and clinical potential of MicroRNAs in irritable bowel syndrome: a concise review, *Dig. Dis. Sci.* 68 (2023) 38–53, <https://doi.org/10.1007/s10620-022-07516-6>.
- [68] X. Chen, Y. Wang, N. Ma, J. Tian, Y. Shao, B. Zhu, Y.K. Wong, Z. Liang, C. Zou, J. Wang, Target identification of natural medicine with chemical proteomics approach: probe synthesis, target fishing and protein identification, *Signal Transduct. Targeted Ther.* 5 (2020) 72, <https://doi.org/10.1038/s41392-020-0186-y>.
- [69] F. Huang, B. Zhang, S. Zhou, X. Zhao, C. Bian, Y. Wei, Chemical proteomics: terra incognita for novel drug target profiling, *Chin. J. Cancer* 31 (2012) 507.
- [70] M. Bantscheff, A. Scholten, A.J.R. Heck, Revealing promiscuous drug–target interactions by chemical proteomics, *Drug Discov. Today* 14 (2009) 1021–1029, <https://doi.org/10.1016/j.drudis.2009.07.001>.
- [71] F. Meissner, J. Geddes-McAlister, M. Mann, M. Bantscheff, The emerging role of mass spectrometry-based proteomics in drug discovery, *Nat. Rev. Drug Discov.* 21 (2022) 637–654, <https://doi.org/10.1038/s41573-022-00409-3>.
- [72] P.J. Hajduk, J. Greer, A decade of fragment-based drug design: strategic advances and lessons learned, *Nat. Rev. Drug Discov.* 6 (2007) 211–219, <https://doi.org/10.1038/nrd2220>.
- [73] E. Barile, M. Pellecchia, NMR-based approaches for the identification and optimization of inhibitors of protein–protein interactions, *Chem. Rev.* 114 (2014) 4749–4763, <https://doi.org/10.1021/cr500043b>.
- [74] A.E. Modell, S.L. Blosser, P.S. Arora, Systematic targeting of protein–protein interactions, *Trends Pharmacol. Sci.* 37 (2016) 702–713, <https://doi.org/10.1016/j.tips.2016.05.008>.
- [75] D.E. Scott, A.R. Bayly, C. Abell, J. Skidmore, Small molecules, big targets: drug discovery faces the protein–protein interaction challenge, *Nat. Rev. Drug Discov.* 15 (2016) 533–550, <https://doi.org/10.1038/nrd.2016.29>.
- [76] J.A. Geraets, K.R. Pothula, G.F. Schröder, Integrating cryo-EM and NMR data, *Curr. Opin. Struct. Biol.* 61 (2020) 173–181, <https://doi.org/10.1016/j.sbi.2020.01.008>.

- [77] M. Pellicchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T.L. James, S.W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe, G. Siegal, Perspectives on NMR in drug discovery: a technique comes of age, *Nat. Rev. Drug Discov.* 7 (2008) 738–745, <https://doi.org/10.1038/nrd2606>.
- [78] C. Schmidt, H. Urlaub, Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies, *Curr. Opin. Struct. Biol.* 46 (2017) 157–168, <https://doi.org/10.1016/j.sbi.2017.10.005>.
- [79] H.-W. Wang, J.-W. Wang, How cryo-electron microscopy and X-ray crystallography complement each other, *Protein Sci.* 26 (2017) 32–39, <https://doi.org/10.1002/pro.3022>.
- [80] J.-P. Renaud, C. Chung, U.H. Danielson, U. Egner, M. Hennig, R.E. Hubbard, H. Nar, Biophysics in drug discovery: impact, challenges and opportunities, *Nat. Rev. Drug Discov.* 15 (2016) 679–698, <https://doi.org/10.1038/nrd.2016.123>.
- [81] I. Shimada, T. Ueda, Y. Kofuku, M.T. Eddy, K. Wüthrich, GPCR drug discovery: integrating solution NMR data with crystal and cryo-EM structures, *Nat. Rev. Drug Discov.* 18 (2019) 59–82, <https://doi.org/10.1038/nrd.2018.180>.
- [82] J.P. Overington, B. Al-Lazikani, A.L. Hopkins, How many drug targets are there? *Nat. Rev. Drug Discov.* 5 (2006) 993–996, <https://doi.org/10.1038/nrd2199>.
- [83] A. Watts, Solid-state NMR in drug design and discovery for membrane-embedded targets, *Nat. Rev. Drug Discov.* 4 (2005) 555–568, <https://doi.org/10.1038/nrd1773>.
- [84] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res.* 34 (2006) D668–D672, <https://doi.org/10.1093/nar/gkj067>.
- [85] Z. Isik, C. Baldow, C.V. Cannistraci, M. Schroeder, Drug target prioritization by perturbed gene expression and network information, *Sci. Rep.* 5 (2015), 17417, <https://doi.org/10.1038/srep17417>.
- [86] L. Rampásek, D. Hidru, P. Smirnov, B. Haibe-Kains, A. Goldenberg, Dr Vae, Improving drug response prediction via modeling of drug perturbation effects, *Bioinformatics* 35 (2019) 3743–3751, <https://doi.org/10.1093/bioinformatics/btz158>.
- [87] H. Itadani, S. Mizuarai, H. Kotani, Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation, *Curr. Genom.* 9 (2008) 349–360.
- [88] J. Cheng, L. Yang, V. Kumar, P. Agarwal, Systematic evaluation of connectivity map for disease indications, *Genome Med.* 6 (2014) 95, <https://doi.org/10.1186/s13073-014-0095-1>.
- [89] P. Smirnov, V. Kofia, A. Maru, M. Freeman, C. Ho, N. El-Hachem, G.-A. Adam, W. Ba-alawi, Z. Safikhani, B. Haibe-Kains, PharmacDB: an integrative database for mining in vitro anticancer drug screening studies, *Nucleic Acids Res.* 46 (2017) D994–D1002, <https://doi.org/10.1093/nar/gkx911>.
- [90] P. Smirnov, Z. Safikhani, N. El-Hachem, D. Wang, A. She, C. Olsen, M. Freeman, H. Selby, D.M.A. Gendoo, P. Grossmann, A.H. Beck, H.J.W.L. Aerts, M. Lupien, A. Goldenberg, B. Haibe-Kains, PharmacGx: an R package for analysis of large pharmacogenomic datasets, *Bioinformatics* 32 (2015) 1244–1246, <https://doi.org/10.1093/bioinformatics/btv723>.
- [91] H. Sharifi-Noghabi, S. Jahangiri-Tazehkand, P. Smirnov, C. Hon, A. Mammoliti, S. K. Nair, A.S. Mer, M. Ester, B. Haibe-Kains, Drug sensitivity prediction from cell line-based pharmacogenomics data: guidelines for developing machine learning models, *Briefings Bioinf.* 22 (2021), <https://doi.org/10.1093/bib/bbab294>.
- [92] A.L. Cohen, R. Soldi, H. Zhang, A.M. Gustafson, R. Wilcox, B.E. Welm, J.T. Chang, E. Johnson, A. Spira, S.S. Jeffrey, A.H. Bild, A pharmacogenomic method for individualized prediction of drug sensitivity, *Mol. Syst. Biol.* 7 (2011) 513, <https://doi.org/10.1038/msb.2011.47>.
- [93] C.C. Pritchard, H.H. Cheng, M. Tewari, MicroRNA profiling: approaches and considerations, *Nat. Rev. Genet.* 13 (2012) 358–369, <https://doi.org/10.1038/nrg3198>.
- [94] F. Feng, B. Shen, X. Mou, Y. Li, H. Li, Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine, *J. Genet. Genom.* 48 (2021) 540–551, <https://doi.org/10.1016/j.jgg.2021.03.007>.
- [95] J. Li, W. Zhao, R. Akbani, W. Liu, Z. Ju, S. Ling, C.P. Vellano, P. Roebuck, Q. Yu, A.K. Eterovic, L.A. Byers, M.A. Davies, W. Deng, Y.N.V. Gopal, G. Chen, E.M. von Euw, D. Slamon, D. Conklin, J. v Heymach, A.F. Gazdar, J.D. Minna, J.N. Myers, Y. Lu, G.B. Mills, H. Liang, Characterization of human cancer cell lines by reverse-phase protein arrays, *Cancer Cell* 31 (2017) 225–239, <https://doi.org/10.1016/j.ccell.2017.01.005>.
- [96] C. Calandrin, J. Drost, Normal and tumor-derived organoids as a drug screening platform for tumor-specific drug vulnerabilities, *STAR Protoc* 3 (2022), 101079, <https://doi.org/10.1016/j.xpro.2021.101079>.
- [97] Y. Lai, X. Wei, S. Lin, L. Qin, L. Cheng, P. Li, Current status and perspectives of patient-derived xenograft models in cancer research, *J. Hematol. Oncol.* 10 (2017) 106, <https://doi.org/10.1186/s13045-017-0470-7>.
- [98] A. Subramanian, R. Narayan, S.M. Corsello, D.D. Peck, T.E. Natoli, X. Lu, J. Gould, J.F. Davis, A.A. Tubelli, J.K. Asiedu, D.L. Lahr, J.E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I.C. Smith, D. Lam, A. Liberson, C. Toder, M. Bagul, M. Orzechowski, O.M. Enache, F. Piccioni, S.A. Johnson, N. J. Lyons, A.H. Berger, A.F. Shamji, A.N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P.A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. v Ronco, J.S. Boehm, S.L. Schreiber, J.G. Doench, J.A. Bittker, D. E. Root, B. Wong, T.R. Golub, A next generation connectivity map: L1000 platform and the first 1,000,000 profiles, *Cell* 171 (2017) 1437–1452.e17, <https://doi.org/10.1016/j.cell.2017.10.049>.
- [99] A.B. Keenan, M.L. Wojciechowicz, Z. Wang, K.M. Jagodnik, S.L. Jenkins, A. Lachmann, A. Ma'ayan, Connectivity mapping: methods and applications, *Annu Rev Biomed Data Sci* 2 (2019) 69–92, <https://doi.org/10.1146/annurev-biodatasci-072018-021211>.
- [100] M. Sirota, J.T. Dudley, J. Kim, A.P. Chiang, A.A. Morgan, A. Sweet-Cordero, J. Sage, A.J. Butte, Discovery and preclinical validation of drug indications using compendia of public gene expression data, *Sci. Transl. Med.* 3 (2011), <https://doi.org/10.1126/scitranslmed.3001318>, 96ra77–96ra77.
- [101] V. Parvathaneni, N.S. Kulkarni, A. Muth, V. Gupta, Drug repurposing: a promising tool to accelerate the drug discovery process, *Drug Discov. Today* 24 (2019) 2076, <https://doi.org/10.1016/j.drudis.2019.06.014>. –2085.
- [102] V.W. Berger, Y. Zhou, Kolmogorov–Smirnov test: overview, in: Wiley StatsRef: Statistics Reference Online, John Wiley & Sons, Ltd, 2014, <https://doi.org/10.1002/9781118445112.stat06558>.
- [103] J. Kim, M. Yoo, J. Kang, A.C. Tan, K-Map, Connecting kinases with therapeutics for drug repurposing and development, *Hum. Genom.* 7 (2013) 20, <https://doi.org/10.1186/1479-7364-7-20>.
- [104] B. Xie, Q. Ding, H. Han, D. Wu, miRCancer: a microRNA–cancer association database constructed by text mining on literature, *Bioinformatics* 29 (2013) 638–644, <https://doi.org/10.1093/bioinformatics/btt014>.
- [105] H. Jiang, C. Hu, M. Chen, The advantages of connectivity map applied in traditional Chinese medicine, *Front. Pharmacol.* 12 (2021), <https://doi.org/10.3389/fphar.2021.474267>.
- [106] M. Yoo, J. Shin, H. Kim, J. Kim, J. Kang, A.C. Tan, Exploring the molecular mechanisms of Traditional Chinese Medicine components using gene expression signatures and connectivity map, *Comput. Methods Progr. Biomed.* 174 (2019) 33–40, <https://doi.org/10.1016/j.cmpb.2018.04.002>.
- [107] F. Araújo, C. Martins, C. Azevedo, B. Sarmiento, Chemical modification of drug molecules as strategy to reduce interactions with mucus, *Adv. Drug Deliv. Rev.* 124 (2018) 98–106, <https://doi.org/10.1016/j.addr.2017.09.020>.
- [108] D.J. Craik, D.J. Adams, Chemical modification of conotoxins to improve stability and activity, *ACS Chem. Biol.* 2 (2007) 457–468, <https://doi.org/10.1021/cb700091j>.
- [109] H.-M. Liu, X.-F. Liu, J.-L. Yao, C.-L. Wang, Y. Yu, R. Wang, Utilization of combined chemical modifications to enhance the blood-brain barrier permeability and pharmacological activity of endomorphin-1, *J. Pharmacol. Exp. Therapeut.* 319 (2006) 308–316, <https://doi.org/10.1124/jpet.106.106484>.
- [110] Y. Lee, X. Hou, J.H. Lee, A. Nayak, V. Alexander, P.K. Sharma, H. Chang, K. Phan, Z.-G. Gao, K.A. Jacobson, S. Choi, L.S. Jeong, Subtle chemical changes cross the boundary between agonist and antagonist: new A3 adenosine receptor homology models and structural network analysis can predict this boundary, *J. Med. Chem.* 64 (2021) 12525–12536, <https://doi.org/10.1021/acs.jmedchem.1c00239>.
- [111] R. Nygaard, T.M. Frimurer, B. Holst, M.M. Rosenkilde, T.W. Schwartz, Ligand binding and micro-switches in 7TM receptor structures, *Trends Pharmacol. Sci.* 30 (2009) 249–259, <https://doi.org/10.1016/j.tips.2009.02.006>.
- [112] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (2015) 436–444, <https://doi.org/10.1038/nature14539>.
- [113] Y. Cao, T.A. Geddes, J.Y.H. Yang, P. Yang, Ensemble deep learning in bioinformatics, *Nat. Mach. Intell.* 2 (2020) 500–508, <https://doi.org/10.1038/s42256-020-0217-y>.
- [114] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proceedings of the 27th International Conference on International Conference on Machine Learning, Omnipress, Madison, WI, USA, 2010, pp. 807–814.
- [115] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, M. Dehmer, An introductory review of deep learning for prediction models with big data, *Front Artif Intell* 3 (2020), <https://doi.org/10.3389/frai.2020.00004>.
- [116] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *J Big Data* 8 (2021) 53, <https://doi.org/10.1186/s40537-021-00444-8>.
- [117] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, *Neural Comput.* 31 (2019) 1235–1270, https://doi.org/10.1162/neco_a_01199.
- [118] I. Ullah, M. Manzo, M. Shah, M.G. Madden, Graph convolutional networks: analysis, improvements and results, *Appl. Intell.* 52 (2022) 9033–9044, <https://doi.org/10.1007/s10489-021-02973-4>.
- [119] G. Eraslan, Z. Avsec, J. Gagneur, F.J. Theis, Deep learning: new computational modelling techniques for genomics, *Nat. Rev. Genet.* 20 (2019) 389–403, <https://doi.org/10.1038/s41576-019-0122-6>.
- [120] Y. LeCun, K. Kavukcuoglu, C. Farabet, Convolutional networks and applications in vision, in: Proceedings of 2010 IEEE International Symposium on Circuits and Systems, 2010, pp. 253–256, <https://doi.org/10.1109/ISCAS.2010.5537907>.
- [121] A. Graves, A. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649, <https://doi.org/10.1109/ICASSP.2013.6638947>, 2013.
- [122] T.N. Kipf, M. Welling, Semi-supervised Classification with Graph Convolutional Networks, 2016. ArXiv Preprint ArXiv:1609.02907.
- [123] A. Dhillion, G.K. Verma, Convolutional neural network: a review of models, methodologies and applications to object detection, *Prog. Artif. Intell.* 9 (2020) 85–112, <https://doi.org/10.1007/s13748-019-00203-0>.
- [124] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, X. Chen, Modulation format recognition and OSNR estimation using CNN-based deep learning, *IEEE Photon. Technol. Lett.* 29 (2017) 1667–1670, <https://doi.org/10.1109/LPT.2017.2742553>.

- [125] R. Nirthika, S. Manivannan, A. Ramanan, R. Wang, Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study, *Neural Comput. Appl.* 34 (2022) 5321–5347, <https://doi.org/10.1007/s00521-022-06953-8>.
- [126] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>, 2016.
- [127] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2017, pp. 2261–2269, <https://doi.org/10.1109/CVPR.2017.243>, 2017.
- [128] J. L.J. Y, K.I.S. Woo Sanghyun, Park, CBAM: convolutional block attention module, in: M. S.C, W.Y. Ferrari Vittorio, Hebert (Eds.), *Computer Vision – ECCV 2018*, Springer International Publishing, Cham, 2018, pp. 3–19.
- [129] D. Nie, X. Wen, Chapter 6 - convolutional neural networks for image synthesis, in: N. Burgos, D. Svoboda (Eds.), *Biomedical Image Synthesis and Simulation*, Academic Press, 2022, pp. 91–104, <https://doi.org/10.1016/B978-0-12-824349-7.00013-X>.
- [130] Y. Han, G. Huang, S. Song, L. Yang, H. Wang, Y. Wang, Dynamic neural networks: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (2022) 7436–7456, <https://doi.org/10.1109/TPAMI.2021.3117837>.
- [131] D. Bahdanau, K. Cho, Y. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, 2014, <https://doi.org/10.48550/ARXIV.1409.0473>. ArXiv Preprint ArXiv:1409.0473.
- [132] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, T.-S. Chua, Sca-Cnn, Spatial and channel-wise attention in convolutional networks for image captioning, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2017, pp. 6298–6306, <https://doi.org/10.1109/CVPR.2017.667>, 2017.
- [133] B. Shah, H. Bhavsar, Time complexity in deep learning models, *Procedia Comput. Sci.* 215 (2022) 202–210, <https://doi.org/10.1016/j.procs.2022.12.023>.
- [134] C. Zhang, P. Benz, D.M. Argaw, S. Lee, J. Kim, F. Rameau, J.-C. Bazin, I.S. Kweon, ResNet or DenseNet? Introducing dense shortcuts to ResNet, in: *2021 IEEE Winter Conference on Applications of Computer Vision, WACV*, 2021, pp. 3549–3558, <https://doi.org/10.1109/WACV48630.2021.00359>.
- [135] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Nature* 323 (1986) 533–536, <https://doi.org/10.1038/323533a0>.
- [136] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L. D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural Comput.* 1 (1989) 541–551, <https://doi.org/10.1162/neco.1989.1.4.541>.
- [137] S. Theodoridis, Chapter 18 - neural networks and deep learning, in: S. Theodoridis (Ed.), *Machine Learning*, second ed., Academic Press, 2020, pp. 901–1038, <https://doi.org/10.1016/B978-0-12-818803-3.00030-1>.
- [138] J. Teuwen, N. Moriaikov, Chapter 20 - convolutional neural networks, in: S. K. Zhou, D. Rueckert, G. Fichtinger (Eds.), *Handbook of Medical Image Computing and Computer Assisted Intervention*, Academic Press, 2020, pp. 481–501, <https://doi.org/10.1016/B978-0-12-816176-0.00025-9>.
- [139] C. Reinders, H. Ackermann, M.Y. Yang, B. Rosenhahn, Chapter 4 - learning convolutional neural networks for object detection with very little training data, in: M.Y. Yang, B. Rosenhahn, V. Murino (Eds.), *Multimodal Scene Understanding*, Academic Press, 2019, pp. 65–100, <https://doi.org/10.1016/B978-0-12-817358-9.00010-X>.
- [140] D. Yu, L. Deng, Deep learning and its applications to signal and information processing [exploratory DSP], *IEEE Signal Process. Mag.* 28 (2011) 145–154, <https://doi.org/10.1109/MSP.2010.939038>.
- [141] J. Raitoharju, Chapter 3 - convolutional neural networks, in: A. Iosifidis, A. Tefas (Eds.), *Deep Learning for Robot Perception and Cognition*, Academic Press, 2022, pp. 35–69, <https://doi.org/10.1016/B978-0-32-385787-1.00008-7>.
- [142] F. Baquero, Transmission as a basic process in microbial biology, *Lwoff Award Prize Lecture, FEMS Microbiol. Rev.* 41 (2017) 816–827, <https://doi.org/10.1093/femsre/fux042>.
- [143] D.L. Cook, F.L. Bookstein, J.H. Gennari, Physical properties of biological entities: an introduction to the ontology of physics for biology, *PLoS One* 6 (2011), e28708, <https://doi.org/10.1371/journal.pone.0028708>.
- [144] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information, *Nat. Commun.* 8 (2017) 573, <https://doi.org/10.1038/s41467-017-00680-8>.
- [145] Y. Yamanishi, M. Kotera, M. Kanehisa, S. Goto, Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework, *Bioinformatics* 26 (2010), <https://doi.org/10.1093/bioinformatics/btq176> i246–i254.
- [146] N. Zong, H. Kim, V. Ngo, O. Harismendy, Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations, *Bioinformatics* 33 (2017) 2337–2344, <https://doi.org/10.1093/bioinformatics/btx160>.
- [147] S.K. Mohamed, V. Nováček, A. Nounu, Discovering protein drug targets using knowledge graph embeddings, *Bioinformatics* 36 (2019) 603–610, <https://doi.org/10.1093/bioinformatics/btz600>.
- [148] D.E. Hostallero, Y. Li, A. Emad, Looking at the BiG picture: incorporating bipartite graphs in drug response prediction, *Bioinformatics* 38 (2022) 3609–3620, <https://doi.org/10.1093/bioinformatics/btac383>.
- [149] C. Wu, R.C. Gudivada, B.J. Aronow, A.G. Jegga, Computational drug repositioning through heterogeneous network clustering, *BMC Syst. Biol.* 7 (2013) S6, <https://doi.org/10.1186/1752-0509-7-S5-S6>.
- [150] R.S. Olayan, H. Ashoor, V.B. Bajic, DDR, Efficient computational method to predict drug–target interactions using graph mining and machine learning approaches, *Bioinformatics* 34 (2017) 1164–1173, <https://doi.org/10.1093/bioinformatics/btx731>.
- [151] J.I.F. Bass, A. Diallo, J. Nelson, J.M. Soto, C.L. Myers, A.J.M. Walhout, Using networks to measure similarity between genes: association index selection, *Nat. Methods* 10 (2013) 1169–1176, <https://doi.org/10.1038/nmeth.2728>.
- [152] S. Oliver, Guilt-by-association goes global, *Nature* 403 (2000) 601–602, <https://doi.org/10.1038/35001165>.
- [153] P. Hu, S.C. Janga, M. Babu, J.J. Díaz-Mejía, G. Butland, W. Yang, O. Pogoutse, X. Guo, S. Phanse, P. Wong, S. Chandran, C. Christopoulos, A. Nazarians-Armavil, N.K. Nasser, G. Musso, M. Ali, N. Nazemof, V. Eroukova, A. Golshani, A. Paccanaro, J.F. Greenblatt, G. Moreno-Hagelsieb, A. Emili, Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins, *PLoS Biol.* 7 (2009), e1000096, <https://doi.org/10.1371/journal.pbio.1000096>.
- [154] T. Doğan, E. Akhan Güzelcan, M. Baumann, A. Koyas, H. Atas, I.R. Baxendale, M. Martin, R. Cetin-Atalay, Protein domain-based prediction of drug/compound–target interactions and experimental validation on LIM kinases, *PLoS Comput. Biol.* 17 (2021), e1009171, <https://doi.org/10.1371/journal.pcbi.1009171>.
- [155] Z.-C. Li, M.-H. Huang, W.-Q. Zhong, Z.-Q. Liu, Y. Xie, Z. Dai, X.-Y. Zou, Identification of drug–target interaction from interactome network with ‘guilt-by-association’ principle and topology features, *Bioinformatics* 32 (2015) 1057–1064, <https://doi.org/10.1093/bioinformatics/btv695>.
- [156] J. You, R.D. McLeod, P. Hu, Predicting drug–target interaction network using deep learning model, *Comput. Biol. Chem.* 80 (2019) 90–101, <https://doi.org/10.1016/j.compbiolchem.2019.03.016>.
- [157] M. Zitnik, M. Agrawal, J. Leskovec, Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 34 (2018) i457–i466, <https://doi.org/10.1093/bioinformatics/bty294>.
- [158] D.-S. Cao, Q.-S. Xu, Y.-Z. Liang, propy: a tool to generate various modes of Chou’s PseAAC, *Bioinformatics* 29 (2013) 960–962, <https://doi.org/10.1093/bioinformatics/btt072>.
- [159] D.-S. Cao, Y.-Z. Liang, J. Yan, G.-S. Tan, Q.-S. Xu, S. Liu, PyDPI: freely available Python package for cheminformatics, bioinformatics, and chemogenomics studies, *J. Chem. Inf. Model.* 53 (2013) 3086–3096, <https://doi.org/10.1021/ci400127q>.
- [160] J. Dong, Z.-J. Yao, L. Zhang, F. Luo, Q. Lin, A.-P. Lu, A.F. Chen, D.-S. Cao, PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions, *J. Cheminf.* 10 (2018) 16, <https://doi.org/10.1186/s13321-018-0270-2>.
- [161] R. Muhammad, S. Ahmed, D. Md Farid, S. Shatabda, A. Sharma, A. Dehzangi, PyFeat: a Python-based effective feature generation tool for DNA, RNA and protein sequences, *Bioinformatics* 35 (2019) 3831–3833, <https://doi.org/10.1093/bioinformatics/btz165>.
- [162] Z. Chen, P. Zhao, F. Li, A. Leier, T.T. Marquez-Lago, Y. Wang, G.I. Webb, A. I. Smith, R.J. Daly, K.-C. Chou, J. Song, iFeature: a Python package and web server for features extraction and selection from protein and peptide sequences, *Bioinformatics* 34 (2018) 2499–2502, <https://doi.org/10.1093/bioinformatics/bty140>.
- [163] Z. Chen, P. Zhao, F. Li, T.T. Marquez-Lago, A. Leier, J. Revote, Y. Zhu, D. R. Powell, T. Akutsu, G.I. Webb, K.-C. Chou, A.I. Smith, R.J. Daly, J. Li, J. Song, iLearn: an integrated platform and meta-learner for feature engineering, machine-learning analysis and modeling of DNA, RNA and protein sequence data, *Briefings Bioinf.* 21 (2019) 1047–1057, <https://doi.org/10.1093/bib/bbz041>.
- [164] Z. Chen, P. Zhao, C. Li, F. Li, D. Xiang, Y.-Z. Chen, T. Akutsu, R.J. Daly, G.I. Webb, Q. Zhao, L. Kurgan, J. Song, iLearnPlus: a comprehensive and automated machine-learning platform for nucleic acid and protein sequence analysis, prediction and visualization, *Nucleic Acids Res.* 49 (2021), <https://doi.org/10.1093/nar/gkab122> e60–e60.
- [165] Z. Chen, X. Liu, P. Zhao, C. Li, Y. Wang, F. Li, T. Akutsu, C. Bain, R.B. Gasser, J. Li, Z. Yang, X. Gao, L. Kurgan, J. Song, iFeatureOmega: an integrative platform for engineering, visualization and analysis of features from molecular sequences, structural and ligand data sets, *Nucleic Acids Res.* 50 (2022) W434, <https://doi.org/10.1093/nar/gkac351> –W447.
- [166] S.A. Wildman, G.M. Crippen, Prediction of physicochemical parameters by atomic contributions, *J. Chem. Inf. Comput. Sci.* 39 (1999) 868–873, <https://doi.org/10.1021/ci9903071>.
- [167] A.P. Bento, A. Hersey, E. Félix, G. Landrum, A. Gaulton, F. Atkinson, L.J. Bellis, M. de Veij, A.R. Leach, An open source chemical structure curation pipeline using RDKit, *J. Cheminf.* 12 (2020) 51, <https://doi.org/10.1186/s13321-020-00456-1>.
- [168] M.D. Disney, A.J. Angelbello, Rational design of small molecules targeting oncogenic noncoding RNAs from sequence, *Acc. Chem. Res.* 49 (2016) 2698–2704, <https://doi.org/10.1021/acs.accounts.6b00326>.
- [169] M.G. Costales, J.L. Childs-Disney, H.S. Haniff, M.D. Disney, How we think about targeting RNA with small molecules, *J. Med. Chem.* 63 (2020) 8880–8900, <https://doi.org/10.1021/acs.jmedchem.9b01927>.
- [170] R. Fan, C. Xiao, X. Wan, W. Cha, Y. Miao, Y. Zhou, C. Qin, T. Cui, F. Su, X. Shan, Small molecules with big roles in microRNA chemical biology and microRNA-targeted therapeutics, *RNA Biol.* 16 (2019) 707–718, <https://doi.org/10.1080/15476286.2019.1593094>.
- [171] S.P. Velagapudi, M.D. Cameron, C.L. Haga, L.H. Rosenberg, M. Lafitte, D. R. Duckett, D.G. Phinney, M.D. Disney, Design of a small molecule against an oncogenic noncoding RNA, *Proc. Natl. Acad. Sci. USA* 113 (2016) 5898–5903, <https://doi.org/10.1073/pnas.1523975113>.
- [172] B.M. Suresh, W. Li, P. Zhang, K.W. Wang, I. Yildirim, C.G. Parker, M.D. Disney, A general fragment-based approach to identify and optimize bioactive ligands

- targeting RNA, *Proc. Natl. Acad. Sci. USA* 117 (2020) 33197–33203, <https://doi.org/10.1073/pnas.2012217117>.
- [173] X. Liu, H.S. Haniff, J.L. Childs-Disney, A. Shuster, H. Aikawa, A. Adibekian, M. D. Disney, Targeted degradation of the oncogenic MicroRNA 17-92 cluster by structure-targeting ligands, *J. Am. Chem. Soc.* 142 (2020) 6970–6982, <https://doi.org/10.1021/jacs.9b13159>.
- [174] X. Chen, C. Zhou, C.-C. Wang, Y. Zhao, Predicting potential small molecule-miRNA associations based on bounded nuclear norm regularization, *Briefings Bioinf.* 22 (2021), <https://doi.org/10.1093/bib/bbab328>.
- [175] C.-C. Wang, C.-C. Zhu, X. Chen, Ensemble of kernel ridge regression-based small molecule-miRNA association prediction in human disease, *Briefings Bioinf.* 23 (2021), <https://doi.org/10.1093/bib/bbab431>.
- [176] F. Liu, L. Peng, G. Tian, J. Yang, H. Chen, Q. Hu, X. Liu, L. Zhou, Identifying small molecule-miRNA associations based on credible negative sample selection and random walk, *Front. Bioeng. Biotechnol.* 8 (2020), <https://doi.org/10.3389/fbioe.2020.00131>.
- [177] N.-N. Guan, Y.-Z. Sun, Z. Ming, J.-Q. Li, X. Chen, Prediction of potential small molecule-associated MicroRNAs using graphlet interaction, *Front. Pharmacol.* 9 (2018), <https://doi.org/10.3389/fphar.2018.01152>.
- [178] J. Luo, C. Shen, Z. Lai, J. Cai, P. Ding, Incorporating clinical, chemical and biological information for predicting small molecule-microRNA associations based on non-negative matrix factorization, *IEEE ACM Trans. Comput. Biol. Bioinf* 18 (2021) 2535–2545, <https://doi.org/10.1109/TCBB.2020.2975780>.
- [179] J. Yin, X. Chen, C.-C. Wang, Y. Zhao, Y.-Z. Sun, Prediction of small molecule-miRNA associations by sparse learning and heterogeneous graph inference, *Mol. Pharm.* 16 (2019) 3157–3166, <https://doi.org/10.1021/acs.molpharmaceut.9b00384>.
- [180] J. Qu, X. Chen, Y.-Z. Sun, J.-Q. Li, Z. Ming, Inferring potential small molecule-miRNA association based on triple layer heterogeneous network, *J. Cheminf.* 10 (2018) 30, <https://doi.org/10.1186/s13321-018-0284-9>.
- [181] J. Qu, X. Chen, Y.-Z. Sun, Y. Zhao, S.-B. Cai, Z. Ming, Z.-H. You, J.-Q. Li, Silico prediction of small molecule-miRNA associations based on the HeteSim algorithm, *Mol. Ther. Nucleic Acids* 14 (2019) 274–286, <https://doi.org/10.1016/j.omtn.2018.12.002>.
- [182] C.-C. Wang, X. Chen, A unified framework for the prediction of small molecule-MicroRNA association based on cross-layer dependency inference on multilayered networks, *J. Chem. Inf. Model.* 59 (2019) 5281–5293, <https://doi.org/10.1021/acs.jcim.9b00667>.
- [183] C. Shen, J. Luo, W. Ouyang, P. Ding, H. Wu, Identification of small molecule-miRNA associations with graph regularization techniques in heterogeneous networks, *J. Chem. Inf. Model.* 60 (2020) 6709–6721, <https://doi.org/10.1021/acs.jcim.0c00975>.
- [184] Y. Lv, S. Wang, F. Meng, L. Yang, Z. Wang, J. Wang, X. Chen, W. Jiang, Y. Li, X. Li, Identifying novel associations between small molecules and miRNAs based on integrated molecular networks, *Bioinformatics* 31 (2015) 3638–3644, <https://doi.org/10.1093/bioinformatics/btv417>.
- [185] Y. Zhao, X. Chen, J. Yin, J. Qu, SNMFSSMA: using symmetric nonnegative matrix factorization and Kronecker regularized least squares to predict potential small molecule-microRNA association, *RNA Biol.* 17 (2020) 281–291, <https://doi.org/10.1080/15476286.2019.1694732>.
- [186] K. Deepthi, A.S. Jereesh, An ensemble approach based on multi-source information to predict drug-MiRNA associations via convolutional neural networks, *IEEE Access* 9 (2021) 38331–38341, <https://doi.org/10.1109/ACCESS.2021.3063885>.
- [187] F. Yu, B. Li, J. Sun, J. Qi, R.L. de Wilde, L.A. la Roche, C. Li, S. Ahmad, W. Shi, X. Li, Z. Chen, PSRR: a web server for predicting the regulation of miRNAs expression by small molecules, *Front. Mol. Biosci.* 9 (2022), <https://doi.org/10.3389/fmolb.2022.817294>.
- [188] L. Peng, Y. Tu, L. Huang, Y. Li, X. Fu, X. Chen, DAESTB: inferring associations of small molecule-miRNA via a scalable tree boosting model based on deep autoencoder, *Briefings Bioinf.* 23 (2022), <https://doi.org/10.1093/bib/bbac478>.
- [189] 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 (2019) 1668–1679, <https://doi.org/10.1021/acs.jcim.9b00129>.
- [190] C. Shen, J. Luo, Z. Lai, P. Ding, Multiview joint learning-based method for identifying small-molecule-associated MiRNAs by integrating pharmacological, genomics, and network knowledge, *J. Chem. Inf. Model.* 60 (2020) 4085–4097, <https://doi.org/10.1021/acs.jcim.0c00244>.
- [191] I. Abdelbaky, H. Tayara, K.T. Chong, Identification of miRNA-small molecule associations by continuous feature representation using auto-encoders, *Pharmaceutics* 14 (2022), <https://doi.org/10.3390/pharmaceutics14010003>.
- [192] S. Jamal, V. Periwal, O. Consortium, V. Scaria, Computational analysis and predictive modeling of small molecule modulators of microRNA, *J. Cheminf.* 4 (2012) 16, <https://doi.org/10.1186/1758-2946-4-16>.
- [193] X. Liu, S. Wang, F. Meng, J. Wang, Y. Zhang, E. Dai, X. Yu, X. Li, W. Jiang, SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression, *Bioinformatics* 29 (2012) 409–411, <https://doi.org/10.1093/bioinformatics/bts698>.
- [194] W. Lou, X. Wang, F. Chen, Y. Chen, B. Jiang, H. Zhang, Sequence based prediction of DNA-binding proteins based on hybrid feature selection using random forest and Gaussian Naïve Bayes, *PLoS One* 9 (2014) e86703, <https://doi.org/10.1371/journal.pone.0086703>.
- [195] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [196] C. Huang, Y. Li, C.C. Loy, X. Tang, Learning deep representation for imbalanced classification, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016, pp. 5375–5384, <https://doi.org/10.1109/CVPR.2016.580>.
- [197] F. Jia, Y. Lei, N. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, *Mech. Syst. Signal Process.* 110 (2018) 349–367, <https://doi.org/10.1016/j.ymssp.2018.03.025>.
- [198] Y. Huang, P. Hu, K.C.C. Chan, Z.-H. You, Graph convolution for predicting associations between miRNA and drug resistance, *Bioinformatics* 36 (2019) 851–858, <https://doi.org/10.1093/bioinformatics/btz621>.
- [199] J. Li, K. Lei, Z. Wu, W. Li, G. Liu, J. Liu, F. Cheng, Y. Tang, Network-based identification of microRNAs as potential pharmacogenomic biomarkers for anticancer drugs, *Oncotarget* 7 (2016) 45584–45596, <https://doi.org/10.18632/oncotarget.10052>.
- [200] J.L. Rukov, N. Shomron, MicroRNA pharmacogenomics: post-transcriptional regulation of drug response, *Trends Mol. Med.* 17 (2011) 412–423, <https://doi.org/10.1016/j.molmed.2011.04.003>.
- [201] J.L. Rukov, R. Wilentzik, I. Jaffe, J. Vinther, N. Shomron, PharmacomiR: linking microRNAs and drug effects, *Briefings Bioinf.* 15 (2013) 648–659, <https://doi.org/10.1093/bib/bbs082>.
- [202] X. Chen, N.-N. Guan, Y.-Z. Sun, J.-Q. Li, J. Qu, MicroRNA-small molecule association identification: from experimental results to computational models, *Briefings Bioinf.* 21 (2018) 47–61, <https://doi.org/10.1093/bib/bby098>.
- [203] X. Zhou, E. Dai, Q. Song, X. Ma, Q. Meng, Y. Jiang, W. Jiang, In silico drug repositioning based on drug-miRNA associations, *Briefings Bioinf.* 21 (2019) 498–510, <https://doi.org/10.1093/bib/bbz012>.
- [204] P. del C. Monroig, L. Chen, S. Zhang, G.A. Calin, Small molecule compounds targeting miRNAs for cancer therapy, *Adv. Drug Deliv. Rev.* 81 (2015) 104–116, <https://doi.org/10.1016/j.addr.2014.09.002>.
- [205] R. Fan, C. Xiao, X. Wan, W. Cha, Y. Miao, Y. Zhou, C. Qin, T. Cui, F. Su, X. Shan, Small molecules with big roles in microRNA chemical biology and microRNA-targeted therapeutics, *RNA Biol.* 16 (2019) 707–718, <https://doi.org/10.1080/15476286.2019.1593094>.
- [206] W. Jiang, Y. Qu, Q. Yang, X. Ma, Q. Meng, J. Xu, X. Liu, S. Wang, D-Inc: a comprehensive database and analytical platform to dissect the modification of drugs on lncRNA expression, *RNA Biol.* 16 (2019) 1586–1591, <https://doi.org/10.1080/15476286.2019.1649584>.
- [207] K. Sachdev, M.K. Gupta, A comprehensive review of feature based methods for drug target interaction prediction, *J. Biomed. Inf.* 93 (2019), 103159, <https://doi.org/10.1016/j.jbi.2019.103159>.
- [208] Y. Wang, S. Zhang, F. Li, Y. Zhou, Y. Zhang, Z. Wang, R. Zhang, J. Zhu, Y. Ren, Y. Tan, C. Qin, Y. Li, X. Li, Y. Chen, F. Zhu, Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics, *Nucleic Acids Res.* 48 (2019), <https://doi.org/10.1093/nar/gkz981>. D1031–D1041.
- [209] N. Hecker, J. Ahmed, J. von Eichborn, M. Dunkel, K. Macha, A. Eckert, M. K. Gilson, P.E. Bourne, R. Preissner, SuperTarget goes quantitative: update on drug-target interactions, *Nucleic Acids Res.* 40 (2011), <https://doi.org/10.1093/nar/gkr912>. D1113–D1117.
- [210] X. Chen, C.C. Yan, X. Zhang, X. Zhang, F. Dai, J. Yin, Y. Zhang, Drug-target interaction prediction: databases, web servers and computational models, *Briefings Bioinf.* 17 (2015) 696–712, <https://doi.org/10.1093/bib/bbv066>.
- [211] S.S. MacKinnon, S.A. Madani Tonekaboni, A. Windemuth, Proteome-scale drug-target interaction predictions: approaches and applications, *Curr Protoc* 1 (2021), e302, <https://doi.org/10.1002/cpz1.302>.
- [212] J.P. Hughes, S. Rees, S.B. Kalindjian, K.L. Philpott, Principles of early drug discovery, *Br. J. Pharmacol.* 162 (2011) 1239–1249, <https://doi.org/10.1111/j.1476-5381.2010.01127.x>.
- [213] I. Lee, J. Keum, H. Namee, DeepConv-Dti: Prediction of drug-target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.* 15 (2019) e1007129, <https://doi.org/10.1371/journal.pcbi.1007129>.
- [214] H. Yu, J. Chen, X. Xu, Y. Li, H. Zhao, Y. Fang, X. Li, W. Zhou, W. Wang, Y. Wang, A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data, *PLoS One* 7 (2012), <https://doi.org/10.1371/journal.pone.0037608> e37608.
- [215] A. Ezzat, M. Wu, X.-L. Li, C.-K. Kwok, Drug-target interaction prediction using ensemble learning and dimensionality reduction, *Methods* 129 (2017) 81–88, <https://doi.org/10.1016/j.ymeth.2017.05.016>.
- [216] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information, *Nat. Commun.* 8 (2017) 573, <https://doi.org/10.1038/s41467-017-00680-8>.
- [217] D. Zhou, Z. Xu, W. Li, X. Xie, S. Peng, MultiDTI: drug-target interaction prediction based on multi-modal representation learning to bridge the gap between new chemical entities and known heterogeneous network, *Bioinformatics* 37 (2021) 4485–4492, <https://doi.org/10.1093/bioinformatics/btab473>.
- [218] K. Huang, C. Xiao, L.M. Glass, J. Sun, MolTrans: molecular Interaction Transformer for drug-target interaction prediction, *Bioinformatics* 37 (2020) 830–836, <https://doi.org/10.1093/bioinformatics/btaa880>.
- [219] H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug-target binding affinity prediction, *Bioinformatics* 34 (2018), <https://doi.org/10.1093/bioinformatics/bty593> i821–i829.
- [220] M. Karimi, D. Wu, Z. Wang, Y. Shen, DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks, *Bioinformatics* 35 (2019) 3329–3338, <https://doi.org/10.1093/bioinformatics/btz111>.

- [221] K. Huang, T. Fu, L.M. Glass, M. Zitnik, C. Xiao, J. Sun, DeepPurpose: a deep learning library for drug–target interaction prediction, *Bioinformatics* 36 (2020) 5545–5547, <https://doi.org/10.1093/bioinformatics/btaa1005>.
- [222] A.S. Rifaioğlu, R. Cetin Atalay, D. Cansen Kahraman, T. Doğan, M. Martin, V. Atalay, MDeepPred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery, *Bioinformatics* 37 (2020) 693–704, <https://doi.org/10.1093/bioinformatics/btaa858>.
- [223] F. Li, Z. Zhang, J. Guan, S. Zhou, Effective drug–target interaction prediction with mutual interaction neural network, *Bioinformatics* 38 (2022) 3582–3589, <https://doi.org/10.1093/bioinformatics/btab377>.
- [224] Q. Zhao, H. Zhao, K. Zheng, J. Wang, HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism, *Bioinformatics* 38 (2021) 655–662, <https://doi.org/10.1093/bioinformatics/btab715>.
- [225] Y. Li, G. Qiao, X. Gao, G. Wang, Supervised graph co-contrastive learning for drug–target interaction prediction, *Bioinformatics* 38 (2022) 2847–2854, <https://doi.org/10.1093/bioinformatics/btac164>.
- [226] T. Hinnerichs, R. Hoehndorf, DTI-Voodoo: machine learning over interaction networks and ontology-based background knowledge predicts drug–target interactions, *Bioinformatics* 37 (2021) 4835–4843, <https://doi.org/10.1093/bioinformatics/btab548>.
- [227] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, GraphDTA: predicting drug–target binding affinity with graph neural networks, *Bioinformatics* 37 (2020) 1140–1147, <https://doi.org/10.1093/bioinformatics/btaa921>.
- [228] Z. Zhang, L. Chen, F. Zhong, D. Wang, J. Jiang, S. Zhang, H. Jiang, M. Zheng, X. Li, Graph neural network approaches for drug–target interactions, *Curr. Opin. Struct. Biol.* 73 (2022), 102327, <https://doi.org/10.1016/j.sbi.2021.102327>.
- [229] T. Doğan, E. Akhan Güzelcan, M. Baumann, A. Koyas, H. Atas, I.R. Baxendale, M. Martin, R. Cetin-Atalay, Protein domain-based prediction of drug/compound–target interactions and experimental validation on LIM kinases, *PLoS Comput. Biol.* 17 (2021), e1009171, <https://doi.org/10.1371/journal.pcbi.1009171>.
- [230] J. Lim, S. Ryu, K. Park, Y.J. Choe, J. Ham, W.Y. Kim, Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation, *J. Chem. Inf. Model.* 59 (2019) 3981–3988, <https://doi.org/10.1021/acs.jcim.9b00387>.
- [231] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, Z. Wei, Drug–target affinity prediction using graph neural network and contact maps, *RSC Adv.* 10 (2020) 20701–20712, <https://doi.org/10.1039/D0RA02297G>.
- [232] Q. Yuan, J. Gao, D. Wu, S. Zhang, H. Mamitsuka, S. Zhu, DrugE-Rank: improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank, *Bioinformatics* 32 (2016), <https://doi.org/10.1093/bioinformatics/btw244> i18–i27.
- [233] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, H. Lu, Deep-learning-based drug–target interaction prediction, *J. Proteome Res.* 16 (2017) 1401–1409, <https://doi.org/10.1021/acs.jproteome.6b00618>.
- [234] Y. Masoudi-Sobhanzadeh, Y. Omid, M. Amanlou, A. Masoudi-Nejad, DrugR+: a comprehensive relational database for drug repurposing, combination therapy, and replacement therapy, *Comput. Biol. Med.* 109 (2019) 254–262, <https://doi.org/10.1016/j.combiomed.2019.05.006>.
- [235] H. Shi, S. Liu, J. Chen, X. Li, Q. Ma, B. Yu, Predicting drug–target interactions using Lasso with random forest based on evolutionary information and chemical structure, *Genomics* 111 (2019) 1839–1852, <https://doi.org/10.1016/j.ygeno.2018.12.007>.
- [236] Y. Ding, J. Tang, F. Guo, Identification of drug–target interactions via dual laplacian regularized least squares with multiple kernel fusion, *Knowl. Base Syst.* 204 (2020), 106254, <https://doi.org/10.1016/j.knosys.2020.106254>.
- [237] S.M. Hasan Mahmud, W. Chen, H. Jahan, B. Dai, S.U. Din, A.M. Dzisoo, DeepACTION: a deep learning-based method for predicting novel drug–target interactions, *Anal. Biochem.* 610 (2020), 113978, <https://doi.org/10.1016/j.ab.2020.113978>.
- [238] A. de Gregorio, L. Häberle, P.A. Fasching, V. Müller, I. Schrader, R. Lorenz, H. Forstbauer, T.W.P. Friedl, E. Bauer, N. de Gregorio, M. Deniz, V. Fink, I. Bekes, U. Andergassen, A. Schneeweiss, H. Tesch, S. Mahner, S.Y. Brucker, J.-U. Blohmer, T.N. Fehm, G. Heinrich, K. Lato, M.W. Beckmann, B. Rack, W. Janni, Gemcitabine as adjuvant chemotherapy in patients with high-risk early breast cancer—results from the randomized phase III SUCCESS-A trial, *Breast Cancer Res.* 22 (2020) 111, <https://doi.org/10.1186/s13058-020-01348-w>.
- [239] Y. Hanashima, E. Sano, K. Sumi, Y. Ozawa, C. Yagi, J. Tatsuoka, S. Yoshimura, S. Yamamuro, T. Ueda, T. Nakayama, others, Antitumor effect of lenalidomide in malignant glioma cell lines, *Oncol. Rep.* 43 (2020) 1580–1590.
- [240] D.D. Von Hoff, T. Ervin, F.P. Arena, E.G. Chiorean, J. Infante, M. Moore, T. Seay, S.A. Tjulandin, W.W. Ma, M.N. Saleh, M. Harris, M. Reni, S. Dowden, D. Laheru, N. Bahary, R.K. Ramanathan, J. Tabernero, M. Hidalgo, D. Goldstein, E. Van Cutsem, X. Wei, J. Iglesias, M.F. Renschler, Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine, *N. Engl. J. Med.* 369 (2013) 1691–1703, <https://doi.org/10.1056/NEJMoa1304369>.
- [241] V. Heinemann, Gemcitabine in metastatic breast cancer, *Expert Rev. Anticancer Ther.* 5 (2005) 429–443, <https://doi.org/10.1586/14737140.5.3.429>.
- [242] J. Zhang, X. Zhu, H. Li, B. Li, L. Sun, T. Xie, T. Zhu, H. Zhou, Z. Ye, Piperine inhibits proliferation of human osteosarcoma cells via G2/M phase arrest and metastasis by suppressing MMP-2/-9 expression, *Int. Immunopharm.* 24 (2015) 50–58, <https://doi.org/10.1016/j.intimp.2014.11.012>.
- [243] J. Souglakos, N. Androulakis, K. Syrigos, A. Polyzos, N. Ziras, A. Athanasiadis, S. Kakolyris, S. Tsousis, C. Kouroussis, L. Vamvakas, A. Kalykaki, G. Samonis, D. Mavroudis, V. Georgoulas, FOLFIRI (folinic acid, 5-fluorouracil, oxaliplatin and irinotecan) vs FOLFIRI (folinic acid, 5-fluorouracil and irinotecan) as first-line treatment in metastatic colorectal cancer (MCC): a multicentre randomised phase III trial from the Hellenic Oncology Research Group (HORG), *Br. J. Cancer* 94 (2006) 798–805, <https://doi.org/10.1038/sj.bjc.6603011>.
- [244] J.M. Schmiedel, S.L. Klemm, Y. Zheng, A. Sahay, N. Blüthgen, D.S. Marks, A. van Oudenaarden, MicroRNA control of protein expression noise, *Science* 348 (2015) 128–132, <https://doi.org/10.1126/science.1260388>.
- [245] M. Ha, V.N. Kim, Regulation of microRNA biogenesis, *Nat. Rev. Mol. Cell Biol.* 15 (2014) 509–524, <https://doi.org/10.1038/nrm3838>.
- [246] C.J. Cheng, R. Bahal, I.A. Babar, Z. Pincus, F. Barrera, C. Liu, A. Svoronos, D. T. Braddock, P.M. Glazer, D.M. Engelman, W.M. Saltzman, F.J. Slack, MicroRNA silencing for cancer therapy targeted to the tumour microenvironment, *Nature* 518 (2015) 107–110, <https://doi.org/10.1038/nature13905>.
- [247] L.F.R. Gebert, I.J. MacRae, Regulation of microRNA function in animals, *Nat. Rev. Mol. Cell Biol.* 20 (2019) 21–37, <https://doi.org/10.1038/s41580-018-0045-7>.
- [248] C. Weinreb, A.J. Riesselman, J.B. Ingraham, T. Gross, C. Sander, D.S. Marks, 3D RNA and functional interactions from evolutionary couplings, *Cell* 165 (2016) 963–975, <https://doi.org/10.1016/j.cell.2016.03.030>.
- [249] T. Treiber, N. Treiber, G. Meister, Regulation of microRNA biogenesis and its crosstalk with other cellular pathways, *Nat. Rev. Mol. Cell Biol.* 20 (2019) 5–20, <https://doi.org/10.1038/s41580-018-0059-1>.
- [250] J. O'Brien, H. Hayder, Y. Zayed, C. Peng, Overview of MicroRNA biogenesis, mechanisms of actions, and circulation, *Front. Endocrinol.* 9 (2018), <https://doi.org/10.3389/fendo.2018.00402>.
- [251] M. Fallor, D. Toso, M. Matsunaga, I. Atanasov, R. Senturia, Y. Chen, Z.H. Zhou, F. Guo, DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures, *RNA* 16 (2010) 1570–1583, <https://doi.org/10.1261/rna.2111310>.
- [252] Y. Zeng, R. Yi, B.R. Cullen, Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha, *EMBO J.* 24 (2005) 138–148, <https://doi.org/10.1038/sj.emboj.7600491>.
- [253] T.-C. Chang, M. Perlea, S. Lee, S.L. Salzberg, J.T. Mendell, Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms, *Genome Res.* 25 (2015) 1401–1409, <https://doi.org/10.1101/gr.193607.115>.
- [254] X. Wang, X. Xu, Z. Ma, Y. Huo, Z. Xiao, Y. Li, Y. Wang, Dynamic mechanisms for pre-miRNA binding and export by Exportin-5, *RNA* 17 (2011) 1511–1528, <https://doi.org/10.1261/rna.2732611>.
- [255] L. Agrana-Tamir, N. Shomron, J. Sperling, R. Sperling, Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome, *Nucleic Acids Res.* 42 (2014) 4640–4651, <https://doi.org/10.1093/nar/gkt1413>.
- [256] B.N. Davis-Dusenbery, A. Hata, Mechanisms of control of microRNA biogenesis, *J. Biochem.* 148 (2010) 381–392, <https://doi.org/10.1093/jb/mvq096>.
- [257] S. Yao, MicroRNA biogenesis and their functions in regulating stem cell potency and differentiation, *Biol. Proced. Online* 18 (2016) 8, <https://doi.org/10.1186/s12575-016-0037-y>.
- [258] B.M. Ryan, A.I. Robles, C.C. Harris, Genetic variation in microRNA networks: the implications for cancer research, *Nat. Rev. Cancer* 10 (2010) 389–402, <https://doi.org/10.1038/nrc2867>.
- [259] E.N. Van Meter, J.A. Onyango, K.A. Teske, A review of currently identified small molecule modulators of microRNA function, *Eur. J. Med. Chem.* 188 (2020), 112008, <https://doi.org/10.1016/j.ejmech.2019.112008>.
- [260] J.F. Armstrong, E. Faccenda, S.D. Harding, A.J. Pawson, C. Southan, J. L. Sharman, B. Campo, D.R. Cavanagh, S.P.H. Alexander, A.P. Davenport, M. Spedding, J.A. Davies, N.C. Iuphar, The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY, *Nucleic Acids Res.* 48 (2020), <https://doi.org/10.1093/nar/gkz951>. D1006–D1021.