

AI-Driven Prediction of ncRNA–Drug Interactions: From Computational Modeling to Precision Therapeutics: A review



Blessing A. Aderibigbe ^{1*}, Vuyolwethu Khwaza ^{1*}

Abstract

Drug discovery, if one reflects on its trajectory, has long been anchored in a protein-centric paradigm—effective, certainly, yet increasingly constrained by the reality that a vast proportion of disease-relevant targets remain pharmacologically inaccessible. In parallel, advances in genomics have quietly reshaped this landscape, revealing non-coding RNAs (ncRNAs) not as passive transcriptional artifacts, but as central regulators of gene expression and disease progression. This review explores the evolving intersection between ncRNA biology and artificial intelligence (AI)-driven computational modeling, focusing on the prediction of ncRNA–drug interactions as a pathway toward precision therapeutics. We examine classical approaches, including structure-based docking, and contrast them with emerging machine learning and deep learning frameworks such as graph convolutional networks and sequence-based architectures. These models, while not without limitations, demonstrate an increasing capacity to infer complex interaction patterns even in the absence of complete structural data. Particular attention is given to data ecosystems, network-based representations, and hybrid modeling strategies that integrate biological, chemical, and transcriptomic information. Yet, the field remains marked by uncertainty—data sparsity, interpretability challenges, and validation gaps persist. Still, there is a cautious optimism.

Significance | Reveals how AI-driven modeling unlocks ncRNA-targeted therapeutics, overcoming protein-centric limitations and enabling precision drug discovery in complex diseases.

*Correspondence.

Department of Chemistry, Faculty of Science and Agriculture, University of Fort Hare, Alice Campus, Alice 5700, South Africa. E-mail: baderibigbe@ufh.ac.za, vkhwaza@ufh.ac.za,

Editor Ahmed J Obaid, Ph.D. And accepted by the Editorial Board Apr 19, 2022 (received for review Feb 25, 2022)

As computational tools become more adaptive and biological insights deepen, AI-driven ncRNA–drug prediction may not merely complement traditional pharmacology but redefine it. **Keywords:** Non-coding RNA (ncRNA); AI-driven drug discovery; ncRNA–drug interactions; Graph neural networks; Precision therapeutics

Keywords

Non-coding RNA (ncRNA); AI-driven drug discovery; ncRNA–drug interactions; Graph neural networks; Precision therapeutics

1. Introduction

The trajectory of modern pharmacology—if one pauses to reflect on it—appears to be shifting in ways that are both subtle and profoundly consequential. For decades, drug discovery has been guided, almost instinctively, by a protein-centric paradigm: identify a disease-associated protein, characterize its structure, and design small molecules capable of modulating its function. This approach, while undeniably successful in many therapeutic domains, has gradually begun to reveal its limitations. Estimates suggest that only a modest fraction—roughly 10–15%—of disease-associated proteins have been effectively targeted by existing drugs,

Author Affiliation: ¹ Department of Chemistry, Faculty of Science and Agriculture, University of Fort Hare, Alice Campus, Alice 5700, South Africa.

Please cite this article as: Aderibigbe, B. A., Khwaza, &. (2021). "AI-Driven Prediction of ncRNA–Drug Interactions: From Computational Modeling to Precision Therapeutics: A review", Bioinfo Chem, 3(1), 1–16, 10735

© 2021 Bioinfo Chem, a publication of Eman Research, USA.
This is an open access article under the CC BY-NC-ND license.
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)
(<https://publishing.emanresearch.org>)

leaving a vast portion of the proteome either inaccessible or, more frustratingly, “undruggable” due to structural constraints such as the absence of suitable ligand-binding pockets (Hopkins & Groom, 2002; Overington et al., 2006).

At the same time, advances in genomics have quietly but decisively expanded our understanding of what constitutes the functional genome. It is now widely recognized that a substantial proportion of the human genome—approaching 70%—is transcribed into RNA, yet only a small subset encodes proteins. The remainder comprises a diverse and increasingly significant class of molecules collectively termed non-coding RNAs (ncRNAs). Initially dismissed as transcriptional noise, these molecules are now understood to play critical regulatory roles across nearly all layers of gene expression (Bartel, 2004). This realization, while perhaps unsurprising in hindsight, has fundamentally altered the conceptual landscape of therapeutic targeting.

Among the various classes of ncRNAs, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs) have emerged as particularly influential. These molecules do not merely accompany gene expression processes; rather, they orchestrate them. miRNAs, for instance, regulate post-transcriptional gene expression by binding to target mRNAs, thereby influencing translation and degradation pathways (Bartel, 2004; Calin & Croce, 2006). lncRNAs, in contrast, exhibit a broader functional repertoire, participating in chromatin remodeling, transcriptional regulation, and scaffolding of protein complexes (Chen et al., 2013). CircRNAs—once considered rare anomalies—have now been recognized as stable, conserved molecules capable of acting as miRNA sponges, thereby modulating gene regulatory networks in a more indirect but equally impactful manner (Hansen et al., 2013; Memczak et al., 2013; Salzman et al., 2012).

The biological significance of these ncRNAs is further underscored by their involvement in disease. Aberrant expression or dysfunction of miRNAs has been linked to cancer progression, neurodegenerative disorders, and metabolic diseases (Calin & Croce, 2006; Jiang et al., 2009). Similarly, lncRNAs and circRNAs have been implicated in pathological processes ranging from tumorigenesis to drug resistance mechanisms (Chen et al., 2013; Hansen et al., 2013). In this context, ncRNAs are not merely biomarkers; they represent a vast, largely

untapped reservoir of therapeutic targets.

Yet, translating this promise into practical drug discovery strategies has proven to be anything but straightforward. The concept of targeting RNA is not entirely new. Indeed, the bacterial ribosome—an RNA-rich complex—has long served as a successful target for antibiotics, demonstrating that RNA structures can, under certain conditions, be selectively modulated by small molecules (Davis, 1987; Poehlsgaard & Douthwaite, 2005). More recently, the development and clinical approval of risdiplam, a small molecule that modifies mRNA splicing in spinal muscular atrophy, has provided compelling evidence that RNA-targeted therapies can achieve clinical efficacy in humans (Ratni et al., 2018).

Despite these advances, several challenges persist. One of the most immediate—and perhaps underappreciated—limitations lies in our incomplete understanding of RNA structure. Unlike proteins, which often adopt well-defined and relatively stable three-dimensional conformations, many RNAs exhibit dynamic, context-dependent structures that are difficult to characterize experimentally (Rouskin et al., 2014; Rivas et al., 2017). Techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, while powerful, are not always feasible for large or flexible RNA molecules. Consequently, a substantial portion of the RNA structural landscape remains unresolved.

This structural ambiguity has direct implications for computational modeling. Traditional structure-based approaches, such as molecular docking, rely heavily on accurate three-dimensional representations of both targets and ligands. When applied to RNA, these methods often struggle with issues related to conformational flexibility, inadequate sampling, and limitations in scoring functions (Ruiz-Carmona et al., 2014; Trott & Olson, 2010). While adaptations of docking algorithms have been developed for nucleic acids, their predictive accuracy remains inconsistent, particularly for complex or poorly characterized RNA targets.

In response to these challenges, the field has increasingly turned toward data-driven methodologies, particularly those grounded in artificial intelligence (AI) and machine learning (ML). These approaches, rather than relying solely on explicit structural information, leverage patterns embedded within large-scale biological and chemical datasets. Early efforts in this direction often employed

similarity-based methods, drawing on curated databases of known RNA–ligand interactions to infer potential binding relationships. Platforms such as Inforna exemplify this strategy, enabling sequence-based design of small molecules targeting structured RNAs (Disney et al., 2016; Velagapudi et al., 2014).

More recently, however, there has been a noticeable shift toward more sophisticated modeling frameworks. Deep learning architectures—particularly graph convolutional networks (GCNs) and attention-based models—have demonstrated remarkable capacity to capture complex, non-linear relationships within high-dimensional data (Kipf & Welling, 2017; Alipanahi et al., 2015). By representing RNA sequences and drug molecules as graphs or embeddings, these models can extract meaningful features without requiring detailed structural annotations. This capability is especially valuable in the context of ncRNAs, where structural data are often sparse or unavailable.

The application of these AI-driven approaches to ncRNA–drug interaction prediction holds considerable promise, particularly in areas such as drug resistance and personalized medicine. miRNAs, for example, can regulate the expression of genes involved in drug metabolism and efficacy, thereby influencing therapeutic outcomes (Rees et al., 2016). CircRNAs, through their interaction with miRNAs, add an additional layer of regulatory complexity that may impact drug sensitivity in ways that are only beginning to be understood (Hansen et al., 2013; Jeck et al., 2013). Accurately modeling these interactions could enable the identification of novel therapeutic targets and inform the design of more effective, individualized treatment strategies.

Nevertheless, it would be premature to suggest that current models have fully overcome the inherent challenges of this domain. Data sparsity remains a significant obstacle, as experimentally validated RNA–drug interaction datasets are limited in both size and diversity. Moreover, issues of model interpretability and generalization persist, raising important questions about the reliability of predictions in novel biological contexts. Addressing these limitations will likely require the integration of heterogeneous data sources—combining sequence information, structural predictions, functional annotations, and clinical data—alongside the development of more robust validation frameworks.

In sum, the field of ncRNA-targeted therapeutics, particularly when viewed through the lens of AI-driven modeling, appears to be at a critical juncture. There is, undeniably, a sense of cautious optimism: the tools are becoming more powerful, the data more abundant, and the biological insights more nuanced. Yet, the path forward is not without uncertainty. It is within this tension—between possibility and limitation—that the present review seeks to situate itself, examining both the progress achieved and the challenges that remain in the prediction of ncRNA–drug interactions.

2. Methodology

2.1 Review Design and Conceptual Framework

This study adopts a narrative review methodology, intended to synthesize the evolving landscape of ncRNA–drug interaction prediction from both computational and biological perspectives. Given the interdisciplinary nature of the field—spanning pharmacology, transcriptomics, and artificial intelligence—a rigid systematic framework was considered less suitable than a flexible, concept-driven approach. The narrative design allows for the integration of diverse evidence types, including algorithmic developments, biological insights, and data infrastructure, enabling a more holistic interpretation of how these domains intersect.

The review was structured to follow a logical progression: beginning with foundational biological understanding of non-coding RNAs, moving through classical computational approaches, and culminating in AI-driven predictive frameworks. This layered structure reflects the organization of the manuscript and aligns with the analytical synthesis presented across Tables 1–4.

2.2 Literature Selection and Inclusion Criteria

Relevant studies were identified through a targeted selection strategy emphasizing foundational, high-impact, and methodologically significant publications. Priority was given to peer-reviewed articles and widely cited works that introduced key computational models, biological discoveries, or database resources relevant to ncRNA–drug interactions.

Biological literature was selected to represent the functional characterization of ncRNAs, including seminal work on microRNAs (Bartel, 2004), disease associations (Calin & Croce, 2006), and regulatory mechanisms

involving lncRNAs and circRNAs (Chen et al., 2013; Salzman et al., 2012). These studies establish the biological rationale for considering ncRNAs as therapeutic targets.

Computational methodologies were selected to reflect both traditional and modern paradigms. Classical structure-based approaches, including molecular docking tools such as AutoDock Vina and rDock, were included to represent early predictive strategies (Trott & Olson, 2010; Ruiz-Carmona et al., 2014). These were complemented by machine learning and deep learning frameworks, including convolutional neural networks for sequence-based prediction (Alipanahi et al., 2015) and graph-based models such as Graph Convolutional Networks (Kipf & Welling, 2017).

2.3 Data Sources and Database Integration

To ensure biological and computational relevance, the review integrates key public databases and repositories that underpin ncRNA–drug interaction research. Foundational datasets such as miRBase and LncRNADisease provide curated information on RNA sequences and disease associations (Griffiths-Jones, 2004; Chen et al., 2013), while platforms like miR2Disease and ncDR extend this by incorporating experimentally validated dysregulation patterns and drug resistance relationships (Jiang et al., 2009; Dai et al., 2017).

High-throughput interaction datasets from starBase v2.0 were included to represent large-scale RNA–RNA and protein–RNA networks (Li et al., 2014), alongside pharmacological databases such as DrugBank for drug-related information (Knox et al., 2018). Expression datasets from GEO further contribute transcriptomic context (Barrett et al., 2012). These resources collectively form the empirical foundation for both classical and AI-driven modeling approaches.

2.4 Analytical Strategy and Thematic Synthesis

The synthesis process followed a multi-dimensional analytical framework, integrating four key domains: (i) data infrastructure, (ii) computational models, (iii) structural and functional RNA characteristics, and (iv) evaluation metrics. These domains correspond to the structured presentation of Tables 1–4 and serve as the conceptual backbone of the review.

Computational models were evaluated in terms of their predictive logic, input requirements, and limitations,

particularly in handling RNA structural complexity. Classical docking methods were assessed for their reliance on structural data and associated constraints (Rouskin et al., 2014; Rivas et al., 2017), while AI-based models were analyzed for their ability to extract latent patterns from sequence and network data (Goodfellow et al., 2014).

Evaluation metrics, including AUC–ROC, RMSE, and Spearman correlation, were incorporated to assess model performance across classification and regression tasks (Artusi et al., 2002; Trott & Olson, 2010). Statistical correction methods such as False Discovery Rate (FDR) were also considered to ensure robustness in large-scale predictions (Benjamini & Hochberg, 1995).

2.5 Methodological Limitations and Scope Considerations

As a narrative review, this methodology inherently involves selective interpretation and does not aim for exhaustive coverage of all available studies. While emphasis was placed on foundational and widely recognized works, emerging models and unpublished data may not be fully represented. Additionally, the rapid evolution of AI-driven approaches introduces a temporal limitation, where newer frameworks may extend beyond the scope of this synthesis. Nevertheless, the methodology is designed to provide a coherent and conceptually grounded overview of the field, balancing breadth with interpretive depth.

3. RNA–Small Molecule Interaction Prediction Frameworks

3.1 Reframing the RNA Landscape in Drug Discovery

For a long time—perhaps longer than we now realize—the architecture of drug discovery was built upon a relatively stable assumption: that proteins, with their structured pockets and catalytic precision, would remain the primary and most tractable therapeutic targets. And to be fair, this assumption worked. It yielded decades of pharmacological progress. Yet, as datasets expanded and genomic technologies matured, a quieter realization began to take shape. The vast majority of the human genome does not encode proteins at all. Instead, it gives rise to an intricate and dynamic population of non-coding RNAs (ncRNAs), molecules that were once dismissed as transcriptional noise but are now recognized as central regulators of cellular function (Djebali et al., 2012; Eddy, 2001).

This shift—from protein dominance to transcriptomic complexity—has not been abrupt, but it has been decisive.

With only about 10–15% of disease-relevant proteins considered druggable (Hopkins & Groom, 2002; Overington et al., 2006), the need to explore alternative molecular targets has become increasingly urgent. ncRNAs, by virtue of their regulatory reach and disease relevance, offer precisely such an alternative. Yet targeting them is not simply an extension of protein-based strategies; it requires, almost unavoidably, a rethinking of the computational frameworks that underpin drug discovery.

3.2 Biological Complexity and Therapeutic Potential of ncRNAs

The appeal of ncRNAs as therapeutic targets lies in their remarkable functional diversity. MicroRNAs (miRNAs), for instance, operate at the post-transcriptional level, binding to complementary sequences on messenger RNAs (mRNAs) and thereby regulating gene expression through degradation or translational repression (Bartel, 2004, 2009). Their dysregulation has been consistently associated with oncogenesis, cardiovascular dysfunction, and neurological disorders (Calin & Croce, 2006; Chang et al., 2009). Long non-coding RNAs (lncRNAs), meanwhile, extend beyond this relatively defined role into a far more heterogeneous functional space. They can act as scaffolds for protein complexes, guides for chromatin modifiers, or even molecular decoys that sequester regulatory factors (Wang & Chang, 2011; Dempsey & Cui, 2017). Their structural plasticity allows them to engage in highly context-dependent interactions, often varying across tissue types and developmental stages. Circular RNAs (circRNAs), perhaps the most structurally distinctive class, introduce an additional layer of regulatory sophistication. Formed through back-splicing events, these covalently closed molecules resist exonuclease degradation, granting them unusual stability within the cellular environment (Memczak et al., 2013; Salzman et al., 2012). Their ability to function as miRNA sponges—effectively modulating entire gene regulatory networks—positions them as both biomarkers and therapeutic targets of considerable interest (Hansen et al., 2013; Jeck et al., 2013). Taken together, these ncRNA classes do not merely supplement protein-based biology; they redefine it. They represent regulatory hubs rather than isolated nodes, and as such, their therapeutic targeting demands a systems-level perspective.

3.3 Limitations of Classical Computational Approaches

If the biological case for targeting ncRNAs is compelling, the computational reality has been, until recently, somewhat less accommodating. Traditional methods such as molecular docking—implemented in tools like AutoDock Vina and rDock—were originally developed with proteins in mind (Ruiz-Carmona et al., 2014; Trott & Olson, 2010). These methods rely on relatively rigid structural assumptions, seeking to optimize ligand orientation within a defined binding pocket. RNA, however, resists such simplification. It is not a static entity but a dynamic ensemble of conformations, often reshaping itself in response to environmental cues or ligand binding (Rouskin et al., 2014). This intrinsic flexibility introduces a level of uncertainty that classical docking approaches struggle to accommodate. Moreover, high-resolution structural data for many ncRNAs—particularly lncRNAs—remain sparse or incomplete, further limiting the applicability of structure-based modeling (Rivas et al., 2017). Even when structural data are available, the scoring functions used in docking algorithms—typically optimized for protein-ligand interactions—fail to capture the unique electrostatic and hydration characteristics of RNA. As a result, predictions are often inconsistent, if not outright misleading. It is, in many ways, a mismatch between method and molecule.

3.4 The Emergence of AI-Driven Frameworks

It is within this context of limitation that artificial intelligence (AI) and machine learning (ML) have emerged—not as incremental improvements, but as fundamentally different approaches. Unlike classical models, AI-driven frameworks do not rely exclusively on predefined structural inputs. Instead, they learn directly from data—sequence patterns, molecular graphs, interaction networks—extracting features that may not be immediately apparent to human observers (Alipanahi et al., 2015; Goodfellow et al., 2014). Early implementations were relatively modest, often based on similarity metrics. The underlying assumption was intuitive: molecules with similar structures or sequences would exhibit similar interaction profiles. Platforms such as Inforna leveraged this principle to design small molecules targeting specific RNA motifs, marking an important early success in sequence-based targeting (Disney et al., 2016; Velagapudi et al., 2014). However, similarity-based methods, while useful, are inherently limited. They struggle to generalize beyond known interaction spaces and are often constrained by the availability of curated

datasets. As the field matured, more sophisticated architectures began to take shape.

3.5 Graph-Based Learning and Network Integration

Graph Convolutional Networks (GCNs) represent one of the most significant advancements in this domain. Biological systems, after all, are naturally graph-like: genes interact with RNAs, RNAs interact with drugs, and these interactions form complex, multi-layered networks. GCNs exploit this structure, learning representations—or embeddings—by aggregating information from neighboring nodes (Kipf & Welling, 2017; Duvenaud et al., 2015). This approach allows for a more holistic understanding of interaction patterns. Rather than evaluating a drug and RNA in isolation, GCNs consider their broader network context, capturing indirect relationships that might otherwise be overlooked. Network integration strategies, such as those proposed by Luo et al. (2017), further enhance this capability by combining heterogeneous data sources—chemical similarity, genomic associations, and functional annotations—into unified predictive models. In practice, this means that predictions are no longer based solely on direct binding evidence but also on inferred relationships, offering a more nuanced and, arguably, more biologically realistic perspective.

3.6 Sequence-Based Deep Learning and Transformer Architectures

Parallel to the development of graph-based methods, another line of innovation has focused on treating biological sequences as a form of language. This perspective—initially inspired by advances in natural language processing—has led to the adaptation of Transformer architectures for biological data (Alipanahi et al., 2015). These models operate by assigning attention weights to different parts of a sequence, effectively learning which regions are most relevant for a given interaction. In the context of ncRNA–drug prediction, this might translate to identifying specific motifs—such as hairpin loops or bulges—that play a critical role in ligand binding. The advantage here is subtle but important. Unlike traditional models, which require explicit feature engineering, Transformer-based frameworks learn features implicitly. They uncover patterns that may not be immediately obvious, capturing long-range dependencies within sequences that simpler models might miss.

3.7 Hybrid and Multi-Modal Learning Strategies

Increasingly, the field is moving toward hybrid frameworks that integrate multiple data modalities. These models combine graph-based representations with sequence-based embeddings, effectively merging structural and contextual information into a single predictive framework. Such integration is not merely a technical refinement; it reflects a deeper understanding of biological complexity. A drug molecule, after all, is both a chemical structure and a sequence of functional groups. Similarly, an RNA molecule is both a sequence and a dynamic structural entity. Multi-modal models attempt to reconcile these dual identities, capturing interactions at multiple scales simultaneously. The result, in many cases, is a marked improvement in predictive accuracy. But perhaps more importantly, these models begin to approximate the multi-dimensional nature of biological systems themselves.

3.8 Data Scarcity and the Role of Augmentation

Despite these advances, one challenge remains persistently difficult: the scarcity of high-quality interaction data. Experimental validation of RNA–drug interactions is both time-consuming and resource-intensive, leading to datasets that are often small and biased toward positive interactions (Baxeavanis, 2011). To address this, researchers have turned to data augmentation and perturbation strategies. By introducing controlled variations into training datasets—modifying RNA sequences, altering molecular structures, or simulating network disruptions—models can be trained to recognize more robust and generalizable patterns. This approach, while not without its limitations, represents a pragmatic solution to an otherwise intractable problem. It allows models to explore the “unknown space” of potential interactions, improving their ability to make predictions in novel contexts.

3.9 Interpretability and the Challenge of Trust

One of the more persistent critiques of AI-driven models—particularly in biomedical applications—is their lack of interpretability. Predictions, however accurate, are of limited utility if their underlying rationale cannot be understood. Recent developments have begun to address this concern. Techniques such as attention visualization and gradient-based mapping allow researchers to identify which features contributed most significantly to a prediction. In practical terms, this might involve highlighting specific nucleotides in an RNA sequence or functional groups within a drug molecule that are critical

Table 1. Foundational Data Repositories Supporting ncRNA and Drug Interaction Research. This table summarizes the major databases and platforms that provide core sequence, disease, expression, target, and pharmacological data used in ncRNA–drug interaction studies. Together, these repositories form the data backbone for computational modeling, biomarker discovery, and therapeutic association analysis.

Resource	Est. Year	RNA Class / Data Type	Primary Data	Curation Type	Clinical Link	Reference	Data Scale
miRBase	2004	miRNA	Sequences, hairpins	Expert curated	Multi-disease relevance	Griffiths-Jones (2004)	>1,900 human hairpins
LncRNADisease	2013	lncRNA	Disease associations	Manual and predicted	Diverse pathologies	Chen et al. (2013)	>160 diseases
miR2Disease	2009	miRNA	Dysregulation patterns	Manual curation	Cancer and neurological disorders	Jiang et al. (2009)	349 ncRNAs
ncDR	2017	ncRNA	Drug resistance associations	Literature mining	Oncology and chemotherapy response	Dai et al. (2017)	5,864 relationships
starBase v2.0	2014	Multi-ncRNA	CLIP-Seq interaction data	High-throughput integration	Gene regulation studies	Li et al. (2014)	2.5 million RNA–RNA links
DrugBank	2006	Small molecules	Pharmacological and drug data	Expert curated	Approved and experimental drugs	Knox et al. (2018)	~10,000 drug entries
GEO	2006	Coding and ncRNA	Raw expression datasets	Community deposition	Experimental and translational studies	Barrett et al. (2012)	>4,000 datasets
miRTarBase	2011	miRNA	Experimentally validated targets	Manual and HTS-supported	mRNA targeting and regulation	Hsu et al. (2011)	>4.4 million interactions
Inforna	2016	Structured RNA	Ligand–motif pairs	Database and design platform	RNA-targeting motif discovery	Disney et al. (2016)	Motif-binding scores

Table 2. Foundational Computational Models and Architectures Used for ncRNA–Small Molecule Interaction Prediction. This table outlines major computational frameworks applied in interaction prediction, ranging from classical physics-based docking tools to deep learning, network-based, and matrix-factorization approaches. It highlights their input requirements, core predictive logic, and principal limitations in RNA-focused applications.

Algorithm / Model	Category	Core Mechanism	Input Format	Primary Outcome	Major Limitation	Reference	Founding Year
AutoDock Vina	Physics-based	Empirical scoring function	3D structural PDB files	Binding pose prediction	Poor handling of RNA flexibility	Trott and Olson (2010)	2010

rDock	Physics-based	Empirical and desolvation scoring	Structural files	Affinity scoring	Difficulty with RNA polyanionic properties	Ruiz-Carmona et al. (2014)	2014
GCN	Deep learning	Neighborhood aggregation	Graph adjacency matrices	Node embeddings	Over-smoothing in deep layers	Kipf and Welling (2017)	2017
DeepBind	Deep learning	CNN-based sequence filters	RNA sequences	Binding specificity score	Lacks structural context	Alipanahi et al. (2015)	2015
Katz Metric	Network-based	Path-counting strategy	Bipartite graphs	Link probability estimation	Computationally intensive	Chen et al. (2015)	2015
Inforna 2.0	Sequence-based	Motif identification	Primary RNA sequence	Bioactive lead identification	Focused mainly on motifs	Disney et al. (2016)	2016
SVD-based Matrix Factorization	Statistical	Matrix decomposition	Interaction matrix	Latent feature extraction	Cold-start problem	Boutsidis et al. (2008)	2008
Label Encoding	Preprocessing	Nominal-to-numeric mapping	Categorical labels	Numerical vector generation	No biological or semantic meaning	Pedregosa et al. (2011)	2011
KNN Imputer	Statistical	Local similarity imputation	Feature vectors	Missing value estimation	Sensitive to distance assumptions	Keerin et al. (2012)	2012

for binding. This level of transparency is not merely desirable; it is essential. It bridges the gap between computational prediction and experimental validation, enabling researchers to refine hypotheses and design more effective therapeutics.

3.10 From Prediction to Clinical Translation

Ultimately, the value of any computational framework lies in its clinical applicability. The approval of RNA-targeted therapies such as risdiplam has demonstrated that small molecules can, indeed, modulate RNA function with therapeutic benefit (Ratni et al., 2018). AI-driven models are now poised to extend this success. By enabling large-scale screening of chemical libraries against ncRNA targets, they offer the potential to identify novel therapeutics and repurpose existing drugs. This is particularly relevant in complex diseases such as cancer, where ncRNAs play a central role in drug resistance and treatment response (Rees et al., 2016). There is, perhaps, a cautious optimism here. The tools are improving, the datasets are expanding, and the biological insights are deepening. Yet challenges remain—data quality, model generalization, and clinical validation among them.

Still, the trajectory is clear. The integration of AI into ncRNA-targeted drug discovery is not merely an incremental step; it represents a fundamental shift in how we conceptualize and approach therapeutic development. And while the full implications of this shift are still unfolding, it seems increasingly likely that the future of pharmacology will be shaped as much by algorithms as by molecules.

4. The Evolving Landscape of RNA–Drug Discovery

4.1 Reconsidering the Non-Coding Transcriptome as a Therapeutic Substrate

It is becoming increasingly difficult—perhaps even intellectually unsustainable—to regard the non-coding transcriptome as a passive byproduct of genomic activity. What once appeared as background noise has, over time, revealed itself to be something far more consequential: a dense, highly regulated, and therapeutically relevant network of RNA species. This conceptual shift, though gradual, has reshaped not only our biological understanding but also the computational strategies we employ to interrogate it.

The transition toward RNA-centric drug discovery does not occur in isolation. Rather, it unfolds alongside a parallel

evolution in data infrastructure, modeling frameworks, and validation paradigms. When considered together—as summarized sequentially in Tables 1–4—these elements form a layered architecture that underpins modern ncRNA–drug interaction prediction. In particular, the progression from curated data repositories (Table 1), through computational models (Table 2), to structural targetability (Table 3), and finally to evaluation frameworks (Table 4), reflects a field that is, in many ways, learning to think differently about both biology and prediction.

4.2 The Foundational Data Ecosystem: Building the ncRNA–Drug Interactome

Any discussion of AI-driven discovery must begin, somewhat unavoidably, with data. Not just its volume, but its structure, reliability, and diversity. The early construction of ncRNA databases laid the groundwork for everything that followed. As outlined in Table 1, resources such as miRBase provided one of the first systematic efforts to catalog microRNA sequences, offering a reference framework for over 1,900 human miRNA hairpins (Griffiths-Jones, 2004). Yet sequence alone, while necessary, proved insufficient. The field gradually pivoted toward functional annotation. Databases like LncRNADisease extended this foundation by linking long non-coding RNAs (lncRNAs) to specific pathological conditions, thereby introducing a clinically meaningful dimension to the data landscape (Chen et al., 2013). Similarly, miR2Disease curated dysregulation patterns associated with cancer and neurological disorders, reinforcing the idea that ncRNAs are not merely structural entities but active participants in disease progression (Jiang et al., 2009).

A more decisive shift occurred with the emergence of specialized repositories such as ncDR, which explicitly mapped ncRNAs to drug resistance mechanisms (Dai et al., 2017). This development—arguably a turning point—allowed computational models to move beyond static associations toward predictive insights about therapeutic response. High-throughput platforms like starBase v2.0 further enriched this ecosystem by integrating CLIP-Seq data, generating millions of RNA–RNA and protein–RNA interactions that could be interpreted as complex biological networks (Li et al., 2014). Taken together, these repositories—alongside pharmacological databases like DrugBank and expression repositories such as GEO—form the empirical backbone of ncRNA–drug modeling (Barrett et al., 2012; Knox et al.,

2018). As summarized in Table 1, their combined contribution lies not only in scale but in the diversity of biological contexts they capture.

4.3 From Docking to Learning: The Evolution of Predictive Architectures

If the data landscape provides the foundation, then computational models define how that foundation is used. Historically, drug discovery relied heavily on physics-based approaches, particularly molecular docking. Tools such as AutoDock Vina and rDock attempted to estimate binding affinity through empirical scoring functions applied to three-dimensional structures (Trott & Olson, 2010; Ruiz-Carmona et al., 2014). These methods, while effective for protein targets, encountered persistent difficulties when applied to RNA. The reasons are not trivial. RNA molecules are structurally dynamic, often adopting multiple conformations that are sensitive to environmental conditions (Rouskin et al., 2014). Moreover, the scarcity of high-resolution structural data—especially for lncRNAs—limits the reliability of docking-based predictions (Rivas et al., 2017).

In response, the field began to explore alternative paradigms, gradually shifting from deterministic modeling toward data-driven learning. As detailed in Table 2, early network-based approaches such as the Katz metric treated interaction prediction as a graph problem, leveraging path-based similarity to infer potential associations (Chen, 2015). While computationally intensive, these models introduced the idea that indirect relationships could be informative. The real transformation, however, emerged with deep learning. Graph Convolutional Networks (GCNs) allowed for the representation of drugs and RNAs as nodes within interconnected systems, enabling the model to learn embeddings that capture both local and global interaction patterns (Kipf & Welling, 2017). This was complemented by sequence-based models such as DeepBind, which demonstrated that convolutional neural networks could extract binding motifs directly from RNA sequences, even in the absence of structural data (Alipanahi et al., 2015).

Platforms like InforNa 2.0 bridged these approaches by integrating sequence-level motif recognition with curated interaction databases, effectively linking computational prediction with experimental feasibility (Disney et al., 2016). As summarized in Table 2, the diversity of these models reflects an ongoing effort to reconcile structural

uncertainty with predictive accuracy.

4.4 Structural Determinants and the Question of Targetability

While computational models continue to evolve, their success ultimately depends on the inherent properties of the targets themselves. Not all RNA molecules are equally amenable to therapeutic intervention. Indeed, as highlighted in Table 3, the structural and functional diversity of ncRNAs introduces varying degrees of targetability. MicroRNAs (miRNAs), for example, are short and relatively transient, making them suitable for antisense inhibition strategies but less accessible to traditional small-molecule binding (Bartel, 2004). Their precursors, however—pre-miRNAs—form stable hairpin structures that provide more defined binding sites, particularly for molecules targeting Dicer processing.

Long non-coding RNAs present a more complex challenge. Their length and structural variability enable a wide range of biological functions, from chromatin remodeling to transcriptional regulation (Wang & Chang, 2011). Yet this same complexity complicates computational prediction, requiring models capable of capturing long-range interactions and tertiary folding patterns. Circular RNAs, by contrast, offer a different kind of opportunity. Their covalently closed structure confers remarkable stability, allowing them to persist in cellular environments where linear RNAs might degrade (Salzman et al., 2012). Their role as miRNA sponges suggests a mechanism through which small molecules could indirectly modulate gene expression networks. In addition to these classes, other RNA elements—such as riboswitches and aptamers—demonstrate that RNA can form highly specific binding pockets, rivaling proteins in their selectivity (Fan et al., 1996; Vicens & Westhof, 2018). As summarized in Table 3, these distinctions are not merely descriptive; they directly inform the choice of computational strategy.

4.5 Evaluating Predictive Performance: Metrics and Benchmarking

Even the most sophisticated model, however, remains speculative without rigorous validation. The field has therefore adopted a suite of evaluation metrics designed to assess predictive performance across different dimensions. As presented in Table 4, these metrics range from classification-based measures to statistical significance tests.

Table 3. Structural and Functional Characteristics of Targetable ncRNA Species and Related RNA Classes. This table compares the major RNA species relevant to therapeutic targeting, with emphasis on their stability, structural features, regulatory roles, and practical targetability. These distinctions are important because RNA class strongly influences both binding feasibility and computational prediction strategy.

RNA Category	Stability	Sequence Length	Primary Structure	Regulatory Mode	Biological Role	Reference	Targetability
miRNA	Low to variable	19–25 nt	Single-stranded	Post-transcriptional regulation	mRNA degradation or repression	Bartel (2004)	High (e.g., antagomirs)
lncRNA	Moderate	>200 nt	Complex tertiary structures	Scaffolding and decoy functions	Epigenetic and transcriptional signaling	Wang and Chang (2011)	Emerging
circRNA	Very high	Variable	Covalently closed circular form	miRNA sponging	Gene regulation	Salzman et al. (2012)	Structure-based
Pre-miRNA	Transient	~70–100 nt	Hairpin loop	Processing intermediate	miRNA maturation signal	Bartel (2004)	Dicer inhibition
Ribosome (rRNA)	High	Long	Complex ribonucleoprotein	Protein synthesis	Catalytic core of translation	Poehlsgaard and Douthwaite (2005)	Established (e.g., antibiotics)
Aptamer	High in vitro	20–60 nt	Defined loop structures	High-affinity molecular recognition	Sensing and targeting	Fan et al. (1996)	Direct binding
Pre-mRNA	Low	Variable	Linear with splice regions	Coding precursor regulation	Splicing target	Ratni et al. (2018)	Splicing modifiers
Repeat elements	High	Repetitive	Structured repeat regions	Translational blockade	Molecular sequestration	Arambula et al. (2009)	Sequence-selective
Riboswitches	Moderate	Variable	Aptamer domain with switching conformation	Conformational regulation	Metabolic control	Vicens et al. (2018)	Metabolite analogs

Table 4. Evaluation Metrics and Benchmarking Parameters Commonly Used in Interaction Prediction Studies. This table summarizes the principal evaluation metrics used to assess ncRNA–drug interaction models, including ranking, classification, regression, and statistical significance measures. The choice of metric depends on the modeling task, the balance of the dataset, and whether prediction aims focus on interaction ranking, affinity estimation, or biomarker discovery.

Metric Name	Logical Basis	Performance Scale	Sensitivity	Primary Objective	Early Implementation	Reference	Focus Area
AUC–ROC	True positive vs. false positive rate	0.5 to 1.0	High for signal discrimination	Distinguish interacting from non-interacting pairs	Graph-based models	Kipf and Welling (2017)	Global ranking

AUPR	Precision vs. recall	0.0 to 1.0	High for imbalanced datasets	Positive-pair prioritization	Link prediction studies	Huang et al. (2017)	Imbalanced data
RMSE	Squared prediction error	0 to higher positive values	Sensitive to outliers	Affinity regression accuracy	Docking and scoring systems	Trott and Olson (2010)	Affinity prediction
Pearson's r	Linear correlation	-1.0 to 1.0	Captures linear pattern agreement	Association strength	Statistical modeling	Artusi et al. (2002)	Trend prediction
Spearman's SCC	Rank correlation	-1.0 to 1.0	Captures monotonic trends	Ranking consistency	Non-linear modeling	Artusi et al. (2002)	Sequence motif ranking
F1-score	Harmonic mean of precision and recall	0.0 to 1.0	Balances false positives and false negatives	Classification balance	ANN and ML models	Pedregosa et al. (2011)	Binary classification
Brier score	Probability calibration error	0.0 to 1.0	Sensitive to calibration quality	Forecast reliability	Probabilistic modeling	Storey (2003)	Risk calibration
Log2 fold change	Fold-difference in expression	Negative to positive values	Sensitive to differential expression magnitude	Expression comparison	Microarray and transcriptomic studies	Benjamini and Hochberg (1995)	Biomarker identification
FDR	Multiple-testing correction	0.0 to 1.0	Controls false discovery	Significance adjustment	Genomic screening	Benjamini and Hochberg (1995)	Statistical power

The Area Under the Receiver Operating Characteristic Curve (AUC–ROC) remains a widely used indicator of global model performance, particularly for distinguishing between interacting and non-interacting pairs (Kipf & Welling, 2017). However, in datasets where positive interactions are relatively rare—a common scenario in biological systems—the Area Under the Precision-Recall Curve (AUPR) provides a more informative measure (Huang et al., 2017). For regression-based tasks, such as predicting binding affinity, metrics like Root Mean Square Error (RMSE) and Spearman rank correlation are essential for evaluating both accuracy and ranking consistency (Artusi et al., 2002; Trott & Olson, 2010). Meanwhile, statistical controls such as the False Discovery Rate (FDR) ensure that identified associations are not artifacts of multiple hypothesis testing (Benjamini & Hochberg, 1995).

Additional preprocessing techniques—including label encoding and KNN imputation—play a supporting role by addressing data sparsity and heterogeneity (Keerin et al., 2012; Pedregosa et al., 2011). As summarized in Table 4, the selection of evaluation metrics is not merely technical; it reflects the underlying objectives of the predictive model.

4.6 From Computational Prediction to Clinical Reality

Perhaps the most compelling validation of these frameworks lies in their clinical translation. The approval of risdiplam for spinal muscular atrophy represents a landmark achievement, demonstrating that small molecules can effectively target RNA splicing mechanisms in vivo (Ratni et al., 2018). This success has broader implications. It suggests that the integration of data curation, AI-driven modeling, structural understanding, and rigorous evaluation—captured collectively in Tables 1 through 4—is not merely theoretical. It is, in fact, actionable.

At the same time, challenges remain. Data incompleteness, model interpretability, and the complexity of biological systems continue to limit predictive accuracy. Yet the trajectory is clear. The field is moving—perhaps cautiously, but steadily—toward a model of precision therapeutics in which ncRNA profiles guide treatment decisions.

5. Limitations

This review, while conceptually comprehensive, is not without limitations. The narrative design, by its nature,

involves selective interpretation rather than exhaustive inclusion, which may lead to the underrepresentation of certain emerging studies or niche methodologies. Additionally, the rapid pace of advancement in AI-driven modeling introduces a temporal constraint—new frameworks, datasets, and validation techniques may already be extending beyond the scope captured here. Another limitation lies in the dependence on available experimental data. Much of the current understanding of ncRNA–drug interactions is shaped by relatively small and sometimes biased datasets, which can influence both the development and evaluation of computational models. Furthermore, challenges related to model interpretability and reproducibility remain unresolved, limiting direct clinical translation. Thus, while the review provides a structured synthesis, it should be interpreted as a reflective snapshot rather than a definitive account of the field.

6. Conclusion

The landscape of drug discovery appears to be undergoing a quiet yet profound transformation. As ncRNAs emerge from the margins of genomic interpretation into central regulatory roles, the need for new predictive frameworks becomes increasingly evident. AI-driven models, despite their current limitations, offer a compelling pathway forward—one that does not depend solely on structural certainty but instead learns from biological complexity itself. Still, the transition is not complete. Questions of data quality, interpretability, and validation persist. Yet, perhaps cautiously, it seems that the convergence of ncRNA biology and computational intelligence may ultimately redefine how therapeutics are discovered and personalized.

Author Contributions

B.A.A. conceptualized the study, designed the review framework, and drafted the original manuscript. V.K. contributed to literature analysis, interpretation of findings, and critically reviewed and edited the manuscript for important intellectual content. All authors read and approved the final version of the manuscript.

References

- Aagaard, L., & Rossi, J. J. (2007). RNAi therapeutics: Principles, prospects and challenges. *Advanced Drug Delivery Reviews*, 59(2-3), 75–86.
- Ahmed, A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., & Smola, A. J. (2013). Distributed large-scale natural graph factorization. In *Proceedings of the 22nd International*

- Conference on World Wide Web (pp. 37–48).
- Alipanahi, B., DeLong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33(8), 831–838.
- Arambula, J. F., Ramisetty, S. R., Baranger, A. M., & Zimmerman, S. C. (2009). A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proceedings of the National Academy of Sciences*, 106(38), 16068–16073.
- Artusi, R., Verderio, P., & Marubini, E. (2002). Bravais-Pearson and Spearman correlation coefficients: Meaning, test of hypothesis and confidence interval. *The International Journal of Biological Markers*, 17(2), 148–151.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefremov, A., & Soboleva, A. (2012). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1), D991–D995.
- Bartel, D. P. (2009). MicroRNAs: Target recognition and regulatory functions. *Cell*, 136(2), 215–233.
- Baxeianis, A. D. (2011). The importance of biological databases in biological discovery. *Current Protocols in Bioinformatics*, 34(1), 1.1.1–1.1.8.
- Belkin, M., Niyogi, P., & Sindhiani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7, 2399–2434.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
- Boutsidis, C., & Gallopoulos, E. (2008). SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4), 1350–1362.
- Calin, G. A., & Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nature Reviews Cancer*, 6(11), 857–871.
- Chang, S., Wen, S., Chen, D., & Jin, P. (2009). Small regulatory RNAs in neurodevelopmental disorders. *Human Molecular Genetics*, 18(R1), R18–R26.
- Chen, G., Wang, Z., Wang, D., Li, C., Liu, M., Chen, X., ... & Cui, Q. (2013). LncRNADisease: A database for long-non-coding RNA-associated diseases. *Nucleic Acids Research*, 41(D1), D983–D986.
- Chen, G., Wang, Z., Wang, D., Li, C., Liu, M., Chen, X., Ma, Y., Cao, C., Sun, Z. J., Yan, Z., Liang, H., Singh, J. P., & Cui, Q. (2013). LncRNADisease: A database for long-non-coding RNA-associated diseases. *Nucleic Acids Research*, 41(D1), D983–D986.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Chen, X. (2015). KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Scientific Reports*, 5(1), 16840.
- Crick, F. (1970). Central Dogma of molecular biology. *Nature*, 227(5258), 561–563.
- Dai, E., Yang, F., Wang, J., Zhou, X., Song, Q., An, W., Wang, L., Wang, Y., & Jiang, Q. (2017). ncDR: A comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics*, 33(24), 4010–4011.
- Davis, B. D. (1987). Mechanism of bactericidal action of aminoglycosides. *Microbiological Reviews*, 51(3), 341–350.
- Dempsey, J. L., & Cui, J. Y. (2017). Long non-coding RNAs: A novel paradigm for toxicology. *Toxicological Sciences*, 155(1), 3–21.
- Disney, M. D., Gallo, S. M., & Velagapudi, S. P. (2016). Inforna 2.0: A platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chemical Biology*, 11(6), 1720–1728.
- Disney, M. D., Winkelsas, A. M., Velagapudi, S. P., Southern, M., Fallahi, M., & Childs-Disney, J. L. (2016). Inforna 2.0: A platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chemical Biology*, 11(6), 1720–1728.
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., ... & Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, 489(7414), 101–108.
- Dobson, C. M. (2004). Chemical space and biology. *Nature*, 432(7019), 824–828.
- Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. *Advances in Neural Information Processing Systems*, 2224–2232.
- Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2(12), 919–929.
- Ellsworth, D. L., Mamula, K. A., Blackburn, H. L., McElean, S., Jellema, G. L., Van Laar, R., ... & Vernalis, M. N. (2014). Intensive cardiovascular risk reduction induces sustainable changes in expression of genes and pathways important to vascular function. *Circulation: Cardiovascular Genetics*, 7(2), 151–160.
- Fan, P., Suri, A. K., Fiala, R., Live, D., & Patel, D. J. (1996). Molecular recognition in the FMN–RNA aptamer complex. *Journal of Molecular Biology*, 258(3), 480–500.
- Fineberg, S. K., Kosik, K. S., & Davidson, B. L. (2009). MicroRNAs potentiate neural development. *Neuron*, 64(3), 303–309.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (pp. 249–256).
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.
- Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Research*, 32(Database issue), D109–D111.
- Guilbert, C., & James, T. L. (2008). Docking to RNA via root-mean-square-deviation-driven

- p>energy minimization with flexible ligands and flexible targets.
- Journal of Chemical Information and Modeling*
- , 48(6), 1257–1268.
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., & Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, 495, 384–388.
- Higgs, P. G., & Lehman, N. (2015). The RNA World: Molecular cooperation at the origins of life. *Nature Reviews Genetics*, 16(1), 7–17.
- Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. *Nature Reviews Drug Discovery*, 1, 727–730.
- Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., ... & Roemer, T. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature*, 526(7575), 672–677.
- Hsu, S. D., Lin, F. M., Wu, W. Y., Liang, C., Huang, N. W., Chan, W. L., Lin, W. T., Chen, G. Z., & Huang, H. D. (2011). miRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research*, 39(Database issue), D163–D169.
- Huang, Y. A., Hu, P., Chan, K. C., & You, Z. H. (2017). Constructing prediction models from expression profiles for large scale lncRNA-miRNA interaction profiling. *Bioinformatics*, 34(5), 812–819.
- Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6), 1239–1249.
- Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., ... & Sharpless, N. E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, 19(2), 141–157.
- Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., & Liu, G. (2009). miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research*, 37(Database issue), D98–D104.
- Keerin, P., Kurutach, W., & Boongoen, T. (2012). Cluster-based KNN missing value imputation for DNA microarray data. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 445–450).
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *ICLR*.
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations (ICLR)*.
- Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., ... & Wishart, D. S. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082.
- Kole, R., Krainer, A. R., & Altman, S. (2012). RNA therapeutics: Beyond RNA interference and antisense oligonucleotides. *Nature Reviews Drug Discovery*, 11(2), 125–140.
- Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778–2786.
- Li, J. H., Liu, S., Zhou, H., Qu, L. H., & Yang, J. H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, 42(D1), D92–D97.
- Lonsdale, J., Thomas, J., Salvatore, M., Rebecca, R., Harris, E., Wright, N., ... & GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585.
- Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., ... & Wang, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, 8(1), 573.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., & Munschauer, M. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495, 333–338.
- Naeem, H., Küffner, R., Csaba, G., & Zimmer, R. (2010). miRsel: Automated extraction of associations between microRNAs and genes from the biomedical literature. *BMC Bioinformatics*, 11(1), 135.
- Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there. *Nature Reviews Drug Discovery*, 5(12), 993–996.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Poehlsgaard, J., & Douthwaite, S. (2005). The bacterial ribosome as a target for antibiotics. *Nature Reviews Microbiology*, 3(11), 870–881.
- Ratni, H., Ebeling, M., Baird, J., Bendels, S., Bylund, J., Chen, K. S., Denk, N., Feng, Z., Green, L., Guerard, M., et al. (2018). Discovery of risdiplam, a selective SMN2 gene splicing modifier. *Journal of Medicinal Chemistry*, 61(15), 6501–6517.
- Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J., Price, E. V., Gill, S., Javaid, S., Coletti, M. E., Jones, V. L., & Bodycombe, N. E. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature Chemical Biology*, 12(2), 109–116.
- Rivas, E., Clements, J., & Eddy, S. R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nature Methods*, 14(1), 45–48.
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., & Weissman, J. S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505(7485), 701–705.
- Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidke, P., Barril, X., Hubbard, R. E., & Morley, S. D. (2014). rDock: A fast, versatile docking program. *PLoS Computational Biology*, 10(4), e1003571.
- Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidke, P., ... & Morley, S. D. (2014). rDock: A fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Computational Biology*, 10(4),

e1003571.

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., & Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform. *PLoS ONE*, 7, e72333.

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., & Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE*, 7(2), e30733.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics*, 31(6), 2013–2035.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461.

Velagapudi, S. P., Cameron, M. D., Haga, C. L., Rosenberg, L. H., Lafitte, M., Duckett, D. R., Phinney, D. G., & Disney, M. D. (2014). Design of small molecules targeting precursor microRNAs. *Nature Chemical Biology*, 10(4), 291–297.

Vicens, Q., & Westhof, E. (2018). Biogenesis, folding, and function of riboswitches. *Methods in Molecular Biology*, 1721, 1–22.

Wang, K. C., & Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular Cell*, 43(6), 904–914.

Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery*, 17(8), 547–558.

Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., ... & Turner, D. H. (1998). Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37(42), 14719–14735.