Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GraPhiT-RBP: Graph Transformers and Residue-Level Cross-Attention for Interpretable RNA–Protein Binding Prediction

Walter Moss

wmoss@iastate.edu

Iowa State University https://orcid.org/0000-0001-6419-5570

Abdelraouf Dapour

Iowa State University

Tuqa Ibrahim

Iowa State University

Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11695206/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

GraPhiT-RBP: Graph Transformers and Residue-Level Cross-Attention for Interpretable RNA–Protein Binding Prediction

Abdelraouf O. Dapour[1], Tuqa Ibrahim[2], Walter N. Moss[1,3,‡].

[1]Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA, United States of America, Ames, IA 50011, USA.

[2]Department of Chemistry, Iowa State University, Ames, IA 50011, USA

[3]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA.

[‡]e-mail: wmoss@iastate.edu

Abstract

Accurate prediction of RNA-protein binding is essential for understanding post-transcriptional regulation, yet existing computational approaches typically aggregate features globally, which limits the spatial resolution required to map specific mechanistic determinants of binding. Here, we introduce GraPhiT-RBP, a protein-aware framework that employs graph-guided cross-attention to explicitly model relationships between RNA nucleotides and individual protein residues. By preserving molecular resolution during interaction learning, GraPhiT-RBP enables direct biological interpretation: it recovers experimentally characterized RNA-binding motifs, maps spatial attention directly to known RNA-binding domains, and predicts the disruptive effects of clinically annotated single-nucleotide variants. Evaluated across 261 transcriptome-wide datasets spanning six human cell lines, this architecture achieved superior predictive accuracy (mean AUC 0.9326) and robust zero-shot generalization to previously unseen proteins compared to

eight established methods. Notably, GraPhiT-RBP extracts these transferable recognition features using 69% fewer trainable parameters than the strongest-performing baseline. Together, GraPhiT-RBP offers a highly generalizable, mechanistically interpretable framework for characterizing molecular recognition and investigating previously unexplored RNA-protein regulatory networks.

Introduction

RNA–protein interactions are fundamental to post-transcriptional gene regulation, shaping the fate and function of cellular RNAs throughout their life cycle. RNA-binding proteins (RBPs) associate with specific RNA targets to coordinate diverse processes, including RNA processing and alternative splicing, intracellular localization, stability and degradation, and translational regulation[1, 2, 3]. Large-scale characterization of the human RNA-binding proteome and transcriptome-wide RBP occupancy has further revealed that RNA–protein interactions form extensive regulatory networks connecting individual RBPs to distinct classes of transcripts and RNA-processing pathways[1, 2, 4]. Importantly, RBP recognition is not determined solely by the presence of short sequence motifs; rather, binding reflects the molecular context in which RNA sequence and structural accessibility are recognized by the interacting protein[1, 5]. Consequently, accurately defining the determinants of RNA–protein recognition is essential for understanding post-transcriptional regulatory mechanisms and how their disruption reshapes cellular gene expression.

Experimental advances have enabled increasingly systematic characterization of RBP binding preferences and transcriptome-wide RNA–protein interaction landscapes. In vitro selection and affinity-based assays have provided detailed measurements of the intrinsic recognition preferences of RBPs. Systematic evolution of ligands by exponential enrichment (SELEX) iteratively selects and amplifies high-affinity RNA ligands from highly diverse sequence libraries[6]. RNAcompete uses a designed RNA pool in a competitive binding assay to systematically profile sequence and structural preferences of individual RBPs[7], whereas RNA Bind-n-Seq (RBNS) combines randomized RNA libraries, multiple protein concentrations, and high-throughput sequencing to quantitatively resolve RBP binding specificity[8]. Complementary in vivo approaches based on ultraviolet crosslinking and immunoprecipitation have enabled transcriptome-wide mapping of RBP occupancy under cellular conditions. HITS-CLIP couples UV-mediated protein–RNA crosslinking with immunoprecipitation and high-throughput sequencing to identify RBP-associated RNA regions on a genome-wide scale[9]. PAR-CLIP incorporates photoactivatable ribonucleoside analogs and exploits characteristic crosslink-induced nucleotide transitions to improve the localization of cellular RNA–protein interaction sites[10]. The precise chemical tracking and enrichment of such photoactivatable analogs have since been further optimized to reduce biochemical biases in resulting sequence libraries[11]. Enhanced CLIP (eCLIP) further introduced optimized experimental controls and computational normalization procedures, enabling more robust and scalable profiling of RBP binding across large collections of human datasets[12].

Despite their transformative impact, experimental assays cannot exhaustively interrogate the enormous number of possible interactions among RBPs, RNA sequences, and

cellular environments. Generating binding maps for additional proteins or biological conditions requires dedicated experiments, specialized sample preparation, sequencing, and downstream processing, imposing substantial demands on time, experimental resources, and cost. Measurements may also reflect protocol-specific biases and the biological context in which they were collected, complicating the direct extrapolation of observed binding profiles to untested proteins, RNA targets, or cellular states. Consequently, a large fraction of the potential RNA–protein interaction landscape remains experimentally unexplored, creating a need for computational approaches that can learn transferable determinants of molecular recognition from existing binding measurements and predict previously uncharacterized interactions.

The expansion of large-scale RBP binding datasets has been accompanied by a progressive evolution of computational models for learning RNA recognition preferences. Early deep learning approaches such as DeepBind demonstrated that convolutional neural networks could directly extract predictive sequence patterns from RNA without predefined motifs [13]. Subsequent methods broadened the description of the RNA substrate by incorporating heterogeneous molecular information. iDeep integrated multiple sequence- and annotation-derived feature types through joint representation learning[14], whereas GraphProt combined RNA sequence with predicted secondary structure in a graph-based framework[15].

PRIESSTESS further emphasized interpretable sequence–structure motifs[16], DMSK subsequently used a multi-view learning framework to jointly analyze distinct RNA representations, including predicted secondary-structure information, for RBP binding-

site prediction [17] . Increasing evidence that RNA structure influences RBP recognition motivated models incorporating experimentally measured in-cell RNA structure[18]. PrismNet combined RNA sequence with in vivo icSHAPE profiles to model condition-dependent binding[19], HDRNet subsequently introduced BERT-derived contextual RNA sequence embeddings, moving beyond conventional one-hot encoding, alongside hierarchical multiscale feature extraction to capture broader nucleotide dependencies[20]. Collectively, these developments substantially enriched the representation of RNA, progressing from local sequence motifs toward contextual, structural, and multimodal descriptions of potential binding substrates. However, the protein itself remained implicit in the prediction task: models were generally trained separately for individual RBP–cell-line datasets, requiring a dedicated predictor for each protein and limiting the direct transfer of learned recognition principles to RBPs without corresponding training data[20, 21].

PaRPI recently addressed this protein-specific modeling constraint by reformulating RBP binding prediction as a protein-aware task. Rather than training an independent predictor for each RBP dataset, PaRPI grouped binding data by cell line, integrating datasets from different experimental protocols and batches to enable a single model to learn binding preferences across multiple RBPs[21]. Building upon advances in contextual RNA representation, PaRPI combined BERT-derived RNA sequence embeddings with experimental and predicted structural information, using graph-based learning and Transformer layers to capture nucleotide relationships. Importantly, it introduced protein sequence information through ESM-2 embeddings, allowing predictions to be conditioned on the identity of the interacting protein and extending the framework to RBPs not

represented during training. This represented an important step toward more generalizable RNA–protein binding prediction.

Despite these advances, PaRPI's architecture presents three important limitations that leave room for further improvement in RNA–protein binding prediction. First, a substantial portion of its architecture is dedicated to RNA representation. Starting from contextual BERT-derived embeddings, which already capture sequence-dependent nucleotide relationships, PaRPI applies several successive processing modules, including GraphSAGE for graph-based feature extraction, an Inception module for multiscale feature learning, CBAM for attention-based feature refinement, and a Transformer for capturing longer-range dependencies[21]. Although these components provide complementary mechanisms for enriching RNA features, their sequential use introduces considerable architectural complexity and computational overhead, with much of the model's capacity devoted to processing RNA independently of its interacting protein. This raises the question of whether similarly informative RNA representations could be obtained through a more compact architecture, allowing more of the model's capacity to focus on the molecular interaction being predicted.

Second, despite this extensive RNA processing, PaRPI applies the deep protein–RNA binding predictor (DPRBP) module to aggregate the learned nucleotide features into a single global RNA representation before integrating protein information. Consequently, the detailed RNA representation is compressed without explicitly considering which nucleotide positions may be relevant to the particular protein being evaluated. This early aggregation limits the opportunity to learn protein-dependent relationships at the

nucleotide level, where distinct RBPs may recognize different sequence motifs or structural features within the same RNA.

Third, protein information is similarly represented at a global level. Although ESM-2 provides rich residue-level sequence embeddings, these are aggregated into a single protein representation before being combined with the RNA features through element-wise multiplication[21]. As a result, the final interaction operation involves two global feature vectors rather than explicitly learning relationships between individual RNA positions and protein residues. While this approach enables protein-aware prediction, it does not directly preserve the contributions of distinct RNA-binding domains or individual amino acids during interaction modeling. Such information could be particularly valuable for identifying protein regions associated with RNA recognition and for learning transferable binding preferences across RBPs with diverse recognition mechanisms, including proteins not encountered during training. Together, these limitations motivate an alternative architecture that represents RNA features efficiently, preserves residue-level protein information, and prioritizes direct RNA–protein interaction modeling before downstream feature integration and classification.

To address these limitations, we introduce GraPhiT-RBP, a protein-aware graph and cross-attention framework for RNA–protein binding prediction, designed to shift the focus from extensive RNA-side processing toward learning interactions between RNA and its protein partner. GraPhiT-RBP combines an efficient graph-based representation of RNA sequence and structural information with residue-level protein embeddings, preserving detailed information from both binding partners. Rather than compressing these

representations independently before their interaction, the model employs cross-attention to learn relationships between individual RNA positions and amino acid residues of the corresponding protein. The resulting interaction-informed features are subsequently integrated through a pyramidal multiscale module, enabling a compact, parameter-efficient architecture centered on RNA–protein recognition. Here, we demonstrate that GraPhiT-RBP outperforms existing approaches across diverse human cell lines while requiring substantially fewer trainable parameters than PaRPI. Beyond conventional benchmarking, we evaluate its ability to generalize to previously unseen RNA sequences and held-out RBPs. We further show that increasing protein diversity during training improves predictive performance, supporting the model's capacity to leverage shared recognition features across proteins. Finally, we explore its biological interpretability through the identification of RBP-associated sequence motifs, analysis of the predicted effects of clinically annotated RNA variants on binding, and examination of attention patterns across known RNA-binding domains. Together, these capabilities support improved RNA–protein binding prediction and enable investigation of the molecular features underlying recognition.

Results

1. GraPhiT-RBP integrates molecular representation, residue-level cross-attention, and multiscale interaction learning

To address the limitations of extensive RNA-side processing and late fusion of global molecular representations, we developed GraPhiT-RBP, a compact framework designed to prioritize direct RNA–protein interaction learning (Fig. 1A). The architecture comprises

three main components: (i) a Molecular Representation Module that efficiently encodes RNA sequence and structural information through a graph Transformer while preserving residue-level protein features derived from ESM3; (ii) a Molecular Interaction Module that uses multi-head cross-attention to learn relationships between individual RNA positions and amino acid residues; and (iii) an Interaction-Aware Multiscale Integration module that progressively integrates the resulting protein-conditioned RNA features before binding classification. Together, these components retain molecular information at nucleotide and residue resolution during interaction learning, while concentrating subsequent feature integration on representations already informed by the paired protein.

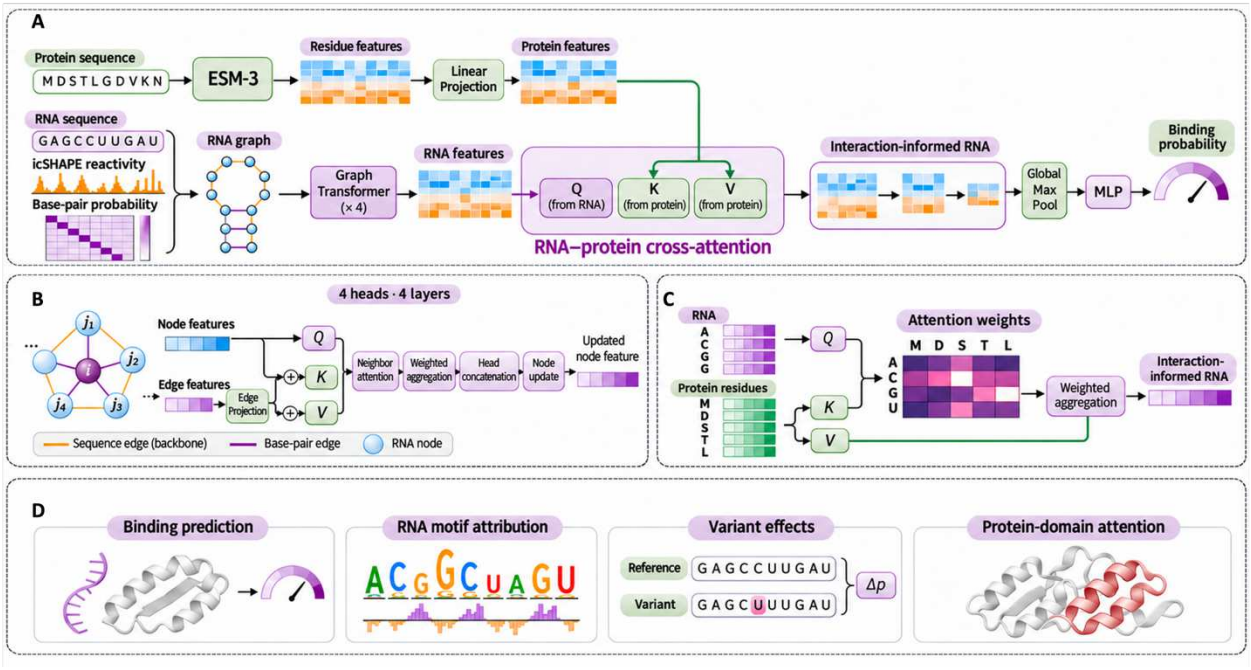


Fig 1. Overview of the GraPhiT-RBP architecture and its applications. (a) Overall architecture of GraPhiT-RBP, integrating RNA sequence, secondary structure, and icSHAPE reactivity with protein residue representations to predict RNA–protein binding probabilities. (b) Graph Transformer module, illustrating how multi-head graph attention integrates RNA node features and edge information to learn structure-aware RNA representations. (c) RNA–protein cross-attention module, illustrating how RNA-derived queries interact with protein-derived keys and values to generate interaction-informed RNA representations. (d) Illustrative applications of GraPhiT-RBP, including RNA–protein

binding prediction, RNA-binding motif attribution, variant effect prediction, and protein-domain attention analysis.

The Molecular Representation Module constructs complementary representations of RNA and protein while preserving the individual molecular positions needed for their interaction (Fig. 1A). On the RNA side, each nucleotide is represented by its sequence identity and experimentally measured *in vivo* icSHAPE reactivity, while the graph incorporates both sequential backbone connections and weighted edges derived from predicted base-pairing probabilities. Rather than relying on a single predicted secondary structure, these probabilistic connections capture alternative pairing relationships within the RNA structural ensemble. Graph Transformer layers propagate information through this combined network, allowing nucleotide features to incorporate context from both adjacent and structurally connected positions (Fig. 1B). This provides a unified way to learn RNA sequence and structural relationships without requiring multiple successive convolutional and self-attention modules. We also assessed whether pretrained BERT-based RNA embeddings provided additional predictive benefits over one-hot nucleotide encoding. Evaluations in Hela and HEK293 showed comparable predictive performance between the two representations. Accordingly, we adopted one-hot encoding in the final architecture, avoiding the additional complexity associated with pretrained RNA embeddings. On the protein side, pretrained ESM3 embeddings provide contextual representations for individual amino acid residues. By retaining these residue-level features rather than compressing the protein into a global vector, the Molecular Representation Module accommodates proteins with different lengths and domain organizations while preserving distinctions among regions potentially involved in RNA recognition.

The Molecular Interaction Module explicitly couples the resulting RNA and protein representations before either partner is reduced to a global summary. Using eight-head cross-attention, nucleotide-level RNA features act as queries, while protein residues provide keys and values. Each RNA position can therefore incorporate information from different residues of its paired protein, producing a protein-conditioned nucleotide representation (Fig. 1C). Unlike the late integration of independently summarized RNA and protein features, this mechanism allows the representation of an RNA position to depend on the particular protein being evaluated. Preserving residue-level protein information also enables the model to learn distinct interaction patterns across proteins with different recognition properties and provides a basis for examining which protein regions receive greater attention during binding prediction.

Finally, the Interaction-Aware Multiscale Integration module consolidates the interaction-conditioned RNA features into a representation suitable for binding classification. Rather than compressing the RNA before it encounters the protein, the Multiscale Integration module operates after cross-attention—when nucleotide representations already contain information from their paired protein residues. A three-stage pyramidal architecture progressively reduces the sequence resolution while applying convolutional feature refinement, integrating interaction information across increasingly broad regions of the RNA. The resulting representation is aggregated and passed to a multilayer perceptron to estimate the probability of RNA–protein binding. By postponing global aggregation until after interaction learning, GraPhiT-RBP combines efficient molecular representation with explicit nucleotide–residue coupling and multiscale integration within a comparatively compact architecture. Beyond binding prediction, GraPhiT-RBP provides a framework for

downstream biological applications, including RNA-binding motif identification, variant effect assessment, and protein-domain attention analysis (Fig. 1D).

2. GraPhiT-RBP achieves broad performance gains across six cell-line benchmarks

To systematically evaluate the predictive performance of GraPhiT-RBP, we compared it with eight previously published RNA–protein binding prediction methods: PaRPI[21], HDRNet[20], PrismNet[19], PRIESSTESS[16], iDeep[14], GraphProt[15], DeepBind[13], and DMSK[17]. The benchmark comprised 261 RBP binding datasets spanning six human cell lines: H9, HEK293, HEK293T, HeLa, HepG2, and K562. Across these datasets, GraPhiT-RBP achieved the highest AUC on 206 of 261 datasets (78.9%), outperforming all competing methods (Fig. 2B). This broad advantage demonstrates that its predictive gains extend across diverse RBPs and cellular contexts.

Among the competing methods, PaRPI and HDRNet provided the strongest overall baselines. Across all 261 datasets, GraPhiT-RBP achieved a mean AUC of 0.9326, compared with 0.9225 for the corrected PaRPI implementation (0.9055 for the originally published results) and approximately 0.8885 for HDRNet, corresponding to relative improvements of 1.1% and 5.0%, respectively. Importantly, GraPhiT-RBP outperformed PaRPI on 220 of 261 datasets (84.3%), demonstrating that its advantage was not restricted to a small number of favorable cases. The overall AUC distributions further supported these improvements, with GraPhiT-RBP showing stronger predictive performance than the competing approaches (Fig. 2A). Paired statistical comparisons confirmed significant improvements over all eight baseline methods (one-sided Wilcoxon

signed-rank tests, all $p < 0.001$). The performance gains were also evident across individual datasets.

The strongest and most consistent improvements were observed in the two largest cell-line benchmarks, HepG2 and K562. GraPhiT-RBP achieved higher AUC than PaRPI on 78 of 84 HepG2 datasets (92.9%) and 110 of 112 K562 datasets (98.2%), with average AUC gains of 1.4 and 1.2 percentage points, respectively. These findings indicate that the benefits of GraPhiT-RBP are particularly consistent across large collections of RBPs, supporting the effectiveness of explicitly modeling RNA–protein interactions across proteins with diverse recognition preferences.

Beyond AUC, GraPhiT-RBP demonstrated substantial improvements in threshold-dependent classification metrics. Compared with PaRPI, GraPhiT-RBP achieved higher accuracy on 238 of 261 datasets (91.2%), higher F1 score on 247 datasets (94.6%), and higher Matthews correlation coefficient (MCC) on 244 datasets (93.5%) (Fig. 2C). Mean accuracy increased from 0.8556 to 0.8787, corresponding to a 2.7% relative improvement, while mean F1 increased from 0.7881 to 0.8174 (3.7% improvement) and mean MCC increased from 0.6840 to 0.7268 (6.3% improvement). These gains were also maintained when compared with the strongest previous method for each dataset and metric, with GraPhiT-RBP achieving the highest accuracy on 227 datasets (87.0%), the highest F1 on 207 datasets (79.3%), and the highest MCC on 233 datasets (89.3%).

Together, these results demonstrate that GraPhiT-RBP provides consistent improvements over established RNA-binding prediction methods across both threshold-independent and threshold-dependent metrics. Notably, these gains were achieved using

a substantially more compact architecture that prioritizes nucleotide–residue interaction modeling, supporting the effectiveness of learning RNA–protein recognition directly rather than relying on extensive independent RNA-side processing.

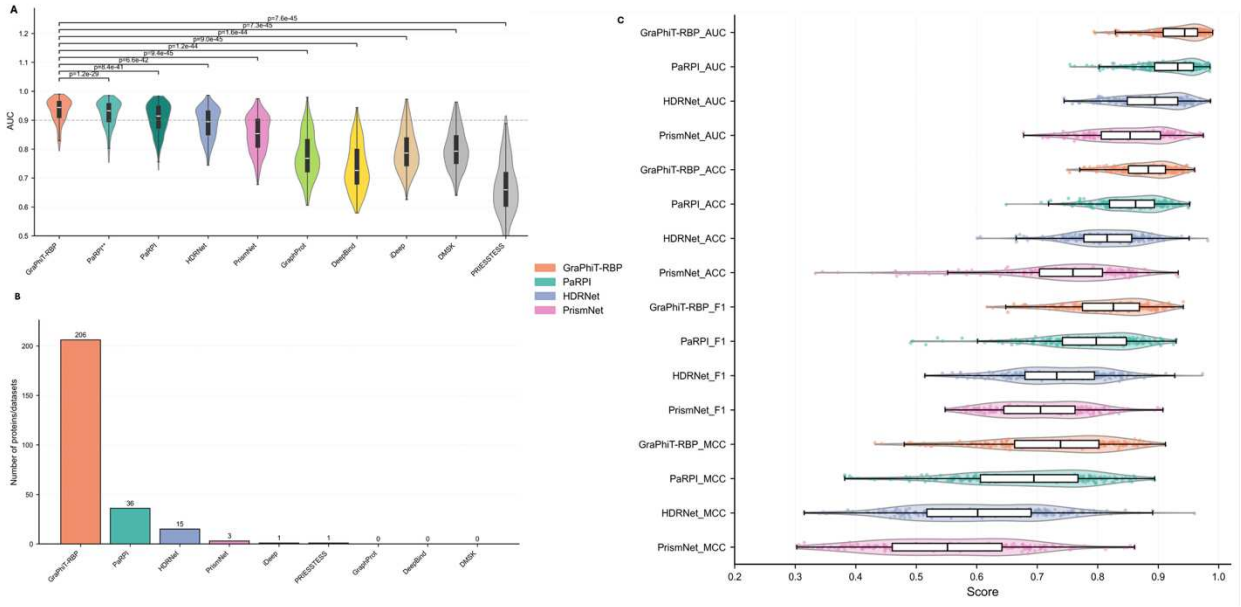


Fig 2. Comparative performance of GraPhiT-RBP across 261 RBP-binding datasets.

(A) Distribution of area under the receiver operating characteristic curve (AUC) values for GraPhiT-RBP and eight previously published prediction methods across 261 RBP-specific datasets from six human cell lines. **(B)** Number of datasets on which each method achieved the highest AUC. GraPhiT-RBP ranked first on 206 of 261 datasets (78.9%). **(C)** Distributions of accuracy, F1 score, and Matthews correlation coefficient (MCC) across the same benchmark datasets. Each observation represents performance calculated separately for an individual RBP dataset. GraPhiT-RBP achieved a mean AUC of 0.9326 and outperformed PaRPI on 220 datasets. Paired comparisons were performed using one-sided Wilcoxon signed-rank tests. PaRPI ** denotes the corrected implementation. Both the original and corrected implementations were evaluated in the initial benchmark; the corrected implementation was used for all subsequent analyses.

3. GraPhiT-RBP outperforms existing methods in RNA sequence-similarity-controlled and held-out protein evaluations

Although GraPhiT-RBP demonstrated strong predictive performance across the standard benchmarks, these evaluations do not fully establish whether the model can generalize beyond molecular sequences represented during training. Similarities between training and test RNA sequences, as well as the inclusion of the same RBPs in both sets, may facilitate prediction without necessarily demonstrating transferable RNA–protein recognition. To investigate these potential issues, we conducted two complementary generalization experiments designed to evaluate performance under more stringent conditions: an RNA sequence-similarity-controlled evaluation and a held-out protein evaluation. Because PaRPI was the only competing method designed to predict binding across multiple RBPs within a unified, protein-aware model, it provides the most appropriate baseline for evaluating generalization to unseen RNA sequences and held-out proteins.

We first evaluated RNA sequence generalization using an 80% sequence-identity threshold to control similarity between training and test RNA sequences (Fig. 3A). This evaluation was performed across all 261 RBP datasets spanning six cell lines, allowing us to assess whether GraPhiT-RBP retained its predictive advantage when highly similar RNA sequences were excluded across the training–test boundary. GraPhiT-RBP achieved a mean AUC of 0.9180, compared with 0.9090 for PaRPI, outperforming PaRPI on 207 of 261 datasets (79.3%). Its advantage also extended beyond AUC, with higher accuracy, F1 score, and MCC on 232 (88.9%), 232 (88.9%), and 232 (88.9%) datasets, respectively. The improvements were particularly consistent in the two largest cell-line benchmarks, where GraPhiT-RBP achieved higher AUC on 77 of 84 HepG2 datasets (91.7%) and 103 of 112 K562 datasets (92.0%). These findings demonstrate that

GraPhiT-RBP maintains its predictive advantage under controlled RNA sequence similarity, supporting its ability to recognize binding patterns beyond closely related RNA sequences encountered during training.

We next investigated whether this advantage extends to RBPs excluded from model training, representing a more challenging setting in which the model must predict binding for previously unseen protein identities (Fig. 3B). We evaluated 22 held-out RBPs in each of the K562 and HepG2 cell lines, comprising 44 protein–cell-line evaluations. Across these evaluations, GraPhiT-RBP achieved a mean AUC of 0.8851**,** compared with 0.8763 for PaRPI, and outperformed PaRPI in 29 of 44 evaluations (65.9%)**.** The improvement was observed in both cell lines, with GraPhiT-RBP achieving higher AUC on 15 of 22 K562 evaluations and 14 of 22 HepG2 evaluations**.** Importantly, the evaluation also included proteins with different levels of sequence similarity to the training proteins. Within the lowest-similarity group (≤30% sequence identity), GraPhiT-RBP outperformed PaRPI on 21 of 33 evaluations (63.6%)**,** indicating that its advantage was not restricted to proteins closely related to those represented during training. Although the magnitude of improvement varied across similarity groups and cell lines, the overall results support the ability of GraPhiT-RBP to transfer learned recognition features to proteins excluded from the training set.

Together, these complementary evaluations demonstrate that GraPhiT-RBP's predictive advantage extends beyond conventional benchmarks to more challenging molecular generalization settings. Its performance under RNA sequence-similarity control and held-out protein evaluation is consistent with the architectural emphasis on retaining

nucleotide- and residue-level representations and learning their interactions before global feature aggregation. Rather than relying exclusively on representations learned for individual RBPs or highly similar RNA sequences, GraPhiT-RBP provides a framework for learning recognition features that can be applied across a broader range of RNA sequences and protein partners.

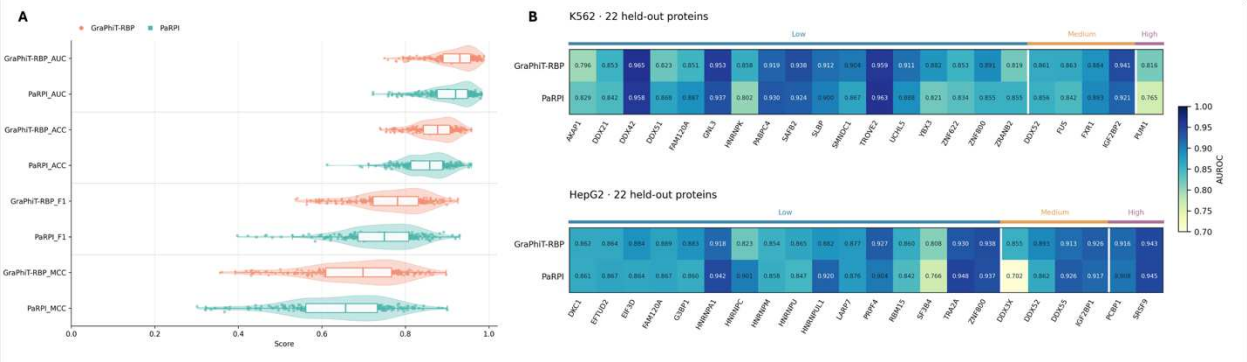


Fig 3. Generalization to RNA sequences and RNA-binding proteins excluded from training.

(A) Comparison of GraPhiT-RBP and PaRPI across 261 RBP datasets after excluding test RNA sequences sharing greater than 80% sequence identity with training sequences. Performance is summarized using AUC, accuracy, F1 score, and MCC. **(B)** Evaluation of protein generalization using 22 completely held-out RBPs from K562 and 22 from HepG2. Each model was trained without binding examples from the corresponding held-out proteins. GraPhiT-RBP achieved a mean AUC of 0.8851 compared with 0.8763 for PaRPI and higher AUC in 29 of 44 protein–cell-line evaluations.

4. GraPhiT-RBP supports cross-cell transfer and benefits from shared learning across diverse RBPs

Although the preceding evaluations demonstrated that GraPhiT-RBP maintains strong predictive performance on unseen RNA sequences and held-out proteins, we further investigated whether the model learns transferable recognition features that can be

shared across cellular environments and diverse RBPs. Specifically, we examined three complementary aspects of model behavior: transfer between cell lines containing common RBPs, predictive performance across different levels of protein sequence similarity, and the effect of incorporating additional RBPs into training.

We first evaluated cross-cell transfer between HepG2 and K562, which share 62 RBPs (Fig. 4A,4B). Although these proteins are common to both cell lines, their binding profiles may differ because of variations in RNA expression, structural accessibility, and the cellular environment. We therefore performed two reciprocal evaluations, training on HepG2 and testing on K562, and then reversing the direction. Both GraPhiT-RBP and PaRPI maintained generally high and comparable AUC values across the shared proteins, indicating that protein-aware prediction can retain useful recognition information across cellular contexts. These results demonstrate that GraPhiT-RBP supports cross-cell transfer even when the underlying RNA-binding data originate from a different cellular environment.

We next investigated whether GraPhiT-RBP's predictive advantage was influenced by protein sequence similarity (Figs. 4B and 4C). Using the MMseqs2-derived classifications from the original PaRPI benchmark, proteins were categorized into low (<30%), medium (30–60%), and high (>60%) sequence-similarity groups. Across all 261 RBP datasets, GraPhiT-RBP achieved higher mean AUC than the competing methods in all three similarity groups, indicating that its performance advantage was not restricted to closely related proteins (Fig. 4C). This pattern was also observed in the K562 subset, where GraPhiT-RBP achieved mean AUCs of 0.9424, 0.9374, and 0.9616 in the low-, medium-

, and high-similarity groups, respectively (Fig. 4D). These values corresponded to relative improvements over PaRPI of 1.3%, 1.1%, and 1.0%, respectively, and improvements over HDRNet of approximately 5.0%, 3.9%, and 3.8%. Notably, the relative advantage over both methods was greatest in the low-similarity group, suggesting that GraPhiT-RBP's benefits extend beyond proteins with closely related sequences. Together, these analyses support its ability to maintain strong predictive performance across RBPs with varying degrees of sequence similarity.

Finally, we investigated whether expanding the number of RBPs available during training improves predictions for proteins already represented in the original training set (Fig. 4E). In HepG2, the training collection was expanded from 62 to 84 RBPs, while in K562 it increased from 90 to 112 RBPs. Importantly, the comparison was performed on the same original protein targets, rather than on the newly added RBPs, allowing us to examine whether additional training proteins contribute useful information to predictions for previously represented targets. Among the 62 original HepG2 RBPs, 56 showed no MMseqs2 sequence-alignment hits against the 22 newly introduced proteins. Of these 56 RBPs, 47 (83.9%) exhibited improved AUC following training-set expansion. Five representative examples are shown in Fig. 4E. Similarly, In K562, 82 of the 90 original RBPs showed no MMseqs2 sequence-alignment hits against any of the 22 newly introduced proteins. Following training-set expansion, 60 of these 82 RBPs (73.2%) exhibited improved AUC. Five representative examples are shown in Fig. 4E. These gains demonstrate that expanding the training collection can improve prediction for existing protein targets, consistent with the model benefiting from recognition information shared across RBPs. Because the expanded collections also contain additional training

examples, these results do not isolate protein diversity from training-data volume as the cause of improvement.

Together, these analyses suggest that GraPhiT-RBP captures recognition features that remain useful across cellular environments and RBPs with different sequence characteristics. Its ability to maintain performance across protein-similarity groups and benefit from expanded protein training sets provides further support for an interaction-centered framework that learns across multiple RNA-binding proteins rather than treating their prediction tasks independently.

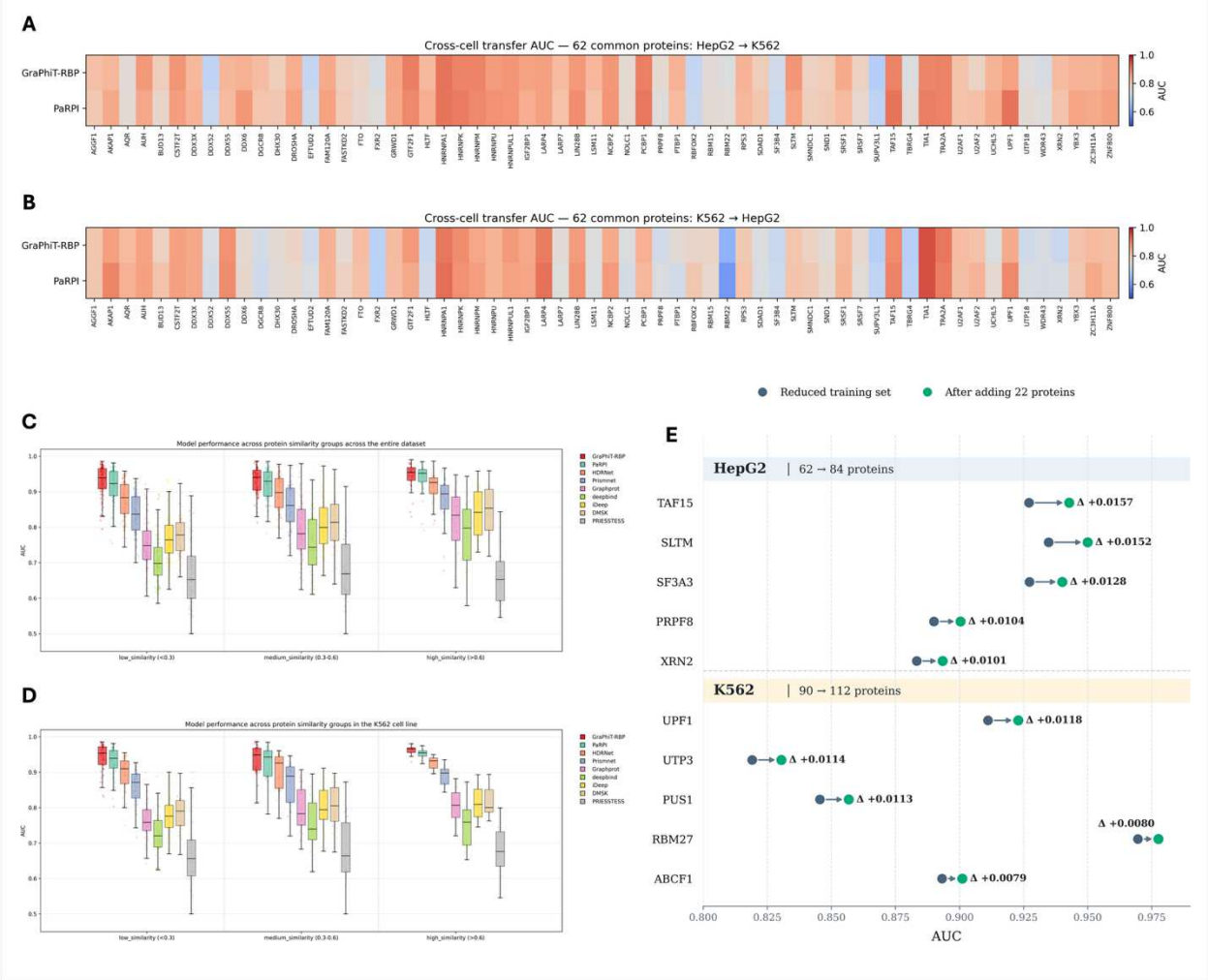


Fig 4. Cross-cell transferability, protein sequence similarity, and expanded RBP training diversity.

(A, B) Reciprocal cross-cell transfer between K562 and HepG2 using 62 RBPs shared by both cell lines, comparing GraPhiT-RBP and PaRPI when trained in one cellular context and evaluated in the other. **(C)** Mean AUC across the complete 261-dataset benchmark stratified by MMseqs2-based protein sequence-similarity groups defined in the PaRPI benchmark: low (less than 30%), medium (30–60%), and high (greater than 60%). **(D)** Corresponding similarity-stratified comparison for the K562 datasets. **(E)** Effect of expanding the training collection from 62 to 84 RBPs in HepG2 and from 90 to 112 RBPs in K562. AUC values are compared on the same original target proteins before and after expansion. Five representative proteins from each cell line are shown, selected among proteins with improved performance and no qualifying MMseqs2 sequence-identity hits against the newly introduced RBPs.

5. Ablation studies demonstrate the importance of molecular representation and interaction-aware integration in GraPhiT-RBP

To investigate the contributions of key architectural components to GraPhiT-RBP's predictive performance, we conducted an ablation study using 17 RBP datasets from the HeLa cell line (Fig. 5A,B). We evaluated four architectural modifications: removal of the RNA graph encoder (ΔRNAGraph), removal of pretrained protein embeddings (ΔProteinEmbedding), replacement of cross-attention with additive fusion (AdditiveFusion), and removal of the interaction-aware pyramidal module (ΔDpRBP). Performance was assessed using protein-macro averaged AUC, accuracy, and F1 score, with the results presented as complementary line and radar plots.

The complete GraPhiT-RBP model achieved the highest performance across all three metrics, with a macro AUC of 0.9356**,** accuracy of 0.8829, and F1 score of 0.8262**.** Removing the RNA graph encoder produced a relatively modest decrease in

performance, reducing AUC to 0.9270, accuracy to 0.8722, and F1 to 0.8095**.** In contrast, removing protein embeddings caused a substantially greater reduction, with AUC decreasing to 0.8861 and F1 to 0.7443, highlighting the importance of residue-level protein information in RNA-binding prediction.

Replacing cross-attention with additive fusion also reduced predictive performance, yielding an AUC of 0.8935, accuracy of 0.8362, and F1 of 0.7541. This supports the value of explicitly modeling nucleotide–residue relationships rather than relying on simpler feature fusion. The largest reduction in AUC was observed when the pyramidal integration module was removed, decreasing AUC to 0.8848, accuracy to 0.8180, and F1 to 0.7401. Together, these results demonstrate that molecular representation, cross-modal interaction learning, and downstream multiscale integration make complementary contributions to the full model's predictive performance.

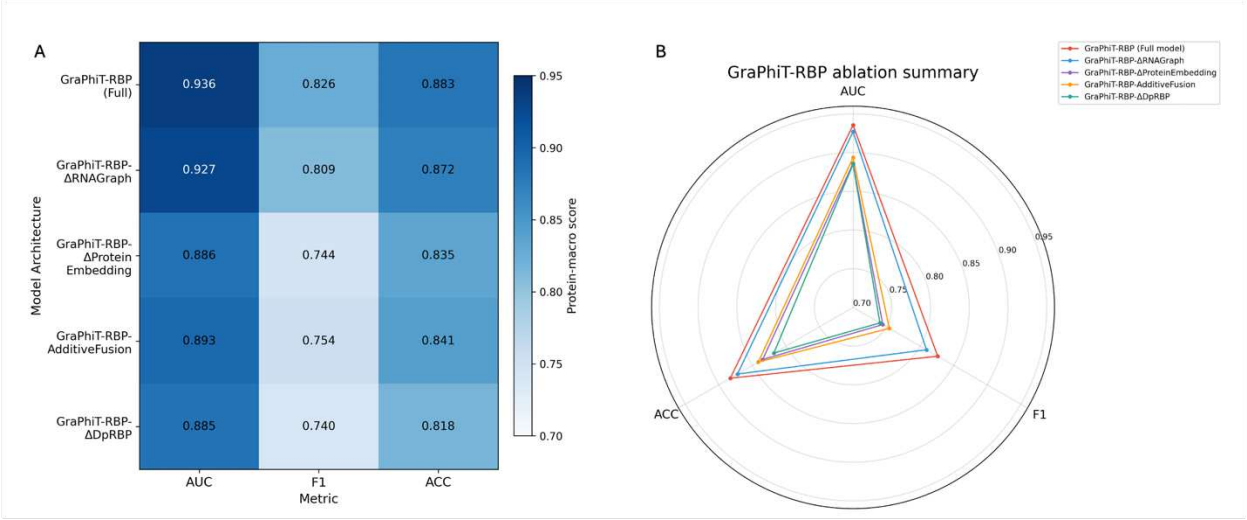


Fig 5. Ablation analysis of key GraPhiT-RBP architectural components.

(**A**) Annotated heatmap of protein-macro scores comparing the full GraPhiT-RBP model against four structural variations (ΔRNAGraph, ΔProteinEmbedding, AdditiveFusion, and ΔDpRBP) across Area Under the Curve (AUC), F1 score, and Accuracy (ACC). Cell color

intensity corresponds to the magnitude of predictive performance. **(B)** Radar plot showing the corresponding performance profiles across the three metrics. The complete model achieved the highest performance across all three metrics, supporting complementary contributions from structural RNA encoding, protein representations, cross-attention, and multiscale feature integration—demonstrating the consistent superiority of the fully integrated model across all evaluation criteria.

6. GraPhiT-RBP learns binding-aware and protein-specific RNA representations

To investigate how GraPhiT-RBP learns RNA–protein recognition features, we visualized the organization of its RNA representations before and after interaction learning (Fig. 6). We first examined CPSF6 as a representative RBP, comparing the initial RNA representation with the final protein-conditioned representation (Fig. 6A). In the initial representation, binding and non-binding RNA samples were broadly intermixed, with limited organization according to binding status. Following RNA–protein interaction learning, positive samples became more concentrated in distinct regions of the representation space, while many negative samples occupied a different region. Although some overlap remained, this change illustrates how incorporating protein information reshapes RNA features into a representation more closely associated with binding status.

We next investigated whether interaction learning also enables GraPhiT-RBP to distinguish the recognition patterns of different RBPs (Fig. 6B). Using five representative proteins—KHDRBS1, CPSF6, GNL3, GTF2F1, and SBDS—we compared their initial RNA representations with those obtained immediately after RNA–protein cross-attention. Before interaction learning, positive RNA samples associated with different RBPs were extensively intermixed, showing little organization according to their protein partners. In contrast, after cross-attention, the representations formed clearly distinguishable protein-associated clusters. This striking reorganization demonstrates that incorporating residue-

level protein information produces RNA representations that differ according to the interacting RBP, allowing the model to distinguish protein-dependent recognition contexts rather than relying solely on RNA features.

Together, these visualization analyses illustrate two complementary properties of GraPhiT-RBP: the transformation of RNA features into binding-associated representations and the emergence of distinct RBP-specific organization following cross-attention. These observations support the intended role of the Molecular Interaction Module in coupling nucleotide-level RNA information with residue-level protein features, providing a qualitative interpretation of how GraPhiT-RBP learns protein-aware binding representations.



Fig 6. Visualization of binding-associated and protein-specific RNA representations.

(A) Two-dimensional visualization of RNA representations for CPSF6 before protein-conditioned interaction learning and after the final interaction-aware representation stage. Positive and negative binding examples are distinguished by color to illustrate their organization in the learned representation space. **(B)** Comparison of initial RNA representations and representations obtained immediately after RNA–protein cross-attention for five representative RBPs: KHDRBS1, CPSF6, GNL3, GTF2F1, and SBDS. Positive examples are colored according to their associated RBP, whereas negative examples are shown in gray. The visualization illustrates the emergence of protein-associated organization following cross-attention. Two-dimensional projections provide qualitative evidence of representation organization rather than direct measurements of binding affinity.

7. GraPhiT-RBP recovers biologically relevant RNA-binding sequence motifs

To investigate whether GraPhiT-RBP learns biologically meaningful RNA sequence preferences beyond accurate binding classification, we examined the nucleotide-level features contributing to its predictions using Integrated Gradients (IG)[22]. This attribution method estimates the contribution of individual input features to a model's prediction by integrating gradients along a path from a reference input to the observed sequence. We applied IG to positively classified RNA sequences associated with RBPs in the K562 cell line, assigning an importance score to each nucleotide based on its contribution to the predicted binding signal. We then extracted six-nucleotide sequence windows (6-mers) from regions with high attribution scores, retaining candidates meeting an 80th-percentile importance threshold. The selected sequences were aggregated to identify recurrent nucleotide patterns and construct representative RNA sequence motifs. Additional details of the attribution procedure, sequence selection, and motif construction are provided in Methods.

To assess whether these computationally identified motifs corresponded to known RNA-binding preferences, we compared the discovered motifs against experimentally derived motif collections from ATtRACT[23] and mCross[24] using TOMTOM[25]. This analysis identified similarities between GraPhiT-RBP-derived motifs and reference motifs associated with known RBPs, providing an independent comparison between sequence patterns highlighted by the model and previously characterized RNA recognition preferences. Representative examples are shown for PUM2, TARDBP, SRSF7, and TRA2A (Fig. 7). Across these proteins, the motifs recovered by GraPhiT-RBP showed recognizable similarities to their corresponding reference motifs in both databases.

For PUM2, the attribution-derived motif recovered the characteristic UGUA sequence preference associated with Pumilio-family recognition. Similarly, the motif identified for TARDBP exhibited a GU-rich pattern consistent with its established RNA-binding preference. The motifs recovered for SRSF7 and TRA2A were enriched in purine-rich sequence patterns, including GA-rich features that resembled their corresponding reference motifs. The agreement between the GraPhiT-RBP-derived motifs and independently characterized binding preferences suggests that nucleotide features contributing to the model's predictions capture biologically relevant aspects of RBP sequence recognition.

Together, these results demonstrate that GraPhiT-RBP can recover recognizable RNA-binding sequence preferences directly from attribution analysis of its predictions. The correspondence between discovered motifs and established reference motifs provides evidence that the model learns informative nucleotide-level recognition patterns, extending its utility beyond binding-site classification toward the interpretation of RBP sequence specificity.

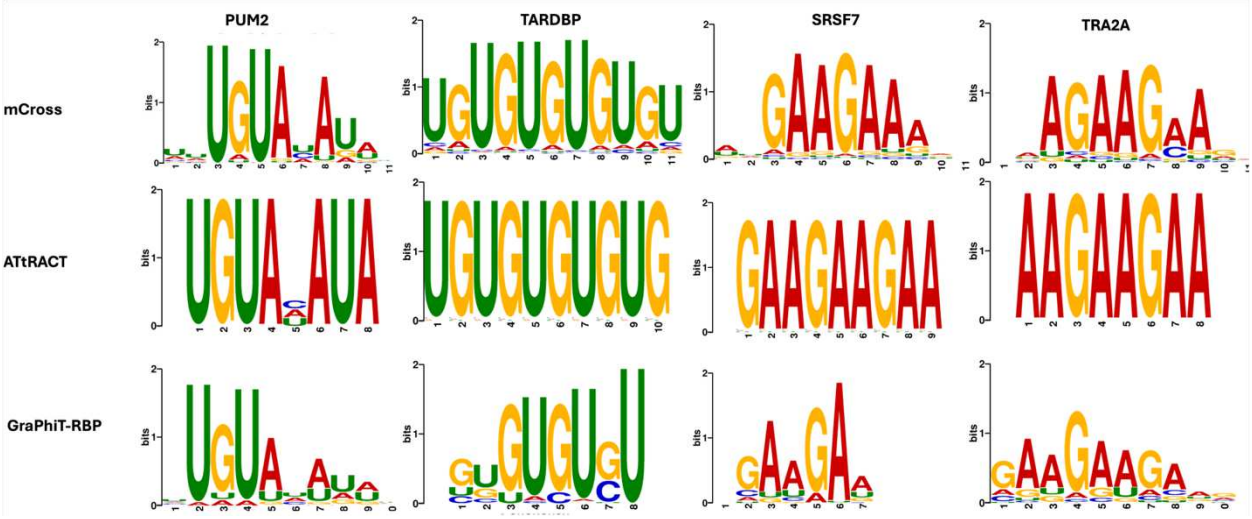


Fig. 7. GraPhiT-RBP identifies RNA sequence motifs consistent with experimentally characterized RBP-binding preferences. Representative sequence logos for PUM2, TARDBP, SRSF7, and TRA2A obtained from mCross (top), ATtRACT (middle), and GraPhiT-RBP attribution-based motif discovery (bottom). GraPhiT-RBP motifs were generated from high-importance RNA sequence regions identified using Integrated Gradients and compared with reference motifs using TOMTOM. The recovered motifs exhibit recognizable similarities to established RBP-binding preferences, including the UGUA-rich motif of PUM2, GU-rich recognition patterns of TARDBP, and purine-rich sequence preferences associated with SRSF7 and TRA2A.

8. GraPhiT-RBP predicts the effects of clinically annotated RNA variants on protein binding

To explore whether GraPhiT-RBP can capture the effects of individual nucleotide changes on RNA–protein recognition, we investigated how annotated clinically significant single-nucleotide variants influence predicted RBP binding. We selected variants from ClinVar[26] and compared the model's predictions for reference RNA sequences and their corresponding mutant sequences, differing by a single nucleotide. For each sequence pair, we calculated the change in predicted binding probability and used Integrated Gradients[22] to examine how the substitution altered nucleotide-level contributions to the prediction. This analysis allowed us to assess whether GraPhiT-RBP responds to small

sequence changes that may affect RBP recognition, while providing a nucleotide-resolved view of the resulting changes in model attribution.

Two representative examples involving PUM2 and SRSF7 are shown in Fig. 8. For PUM2, an A-to-G substitution at position 44 (ClinVar 1342701) reduced the predicted binding probability from 0.989 to 0.103, corresponding to a decrease of 0.886 (Fig. 8A). The substitution also changed the local IG attribution profile, with a pronounced reduction in attribution at the mutated position and redistribution of importance across neighboring nucleotides. Notably, the affected region overlaps a UGUA-containing sequence pattern associated with PUM2 recognition, suggesting that the predicted loss of binding may be related to disruption of sequence features important to the model's recognition of this RNA target.

A similar effect was observed for SRSF7, where an A-to-U substitution at position 49 (ClinVar 982562) reduced predicted binding probability from 0.933 to 0.285, a decrease of 0.647 (Fig. 8B). The variant was accompanied by a shift in the IG attribution profile around the substitution, including a change in the contribution assigned to the mutated nucleotide. These changes indicate that the model's predicted binding response is sensitive to the local sequence alteration and its effects on the surrounding attribution pattern, rather than reflecting an unchanged representation with only a different output probability.

Together, these examples illustrate how GraPhiT-RBP can be applied to evaluate potential changes in RBP recognition associated with single-nucleotide variants. By combining reference-versus-mutant binding predictions with nucleotide-level attribution,

the framework provides a way to prioritize variants for further investigation and identify sequence regions potentially contributing to altered binding. While ΔP captures shifts in model classification confidence rather than physical thermodynamic constants (ΔΔG or Kd), it offers an effective computational screen for sequence substitutions that perturb canonical recognition features.

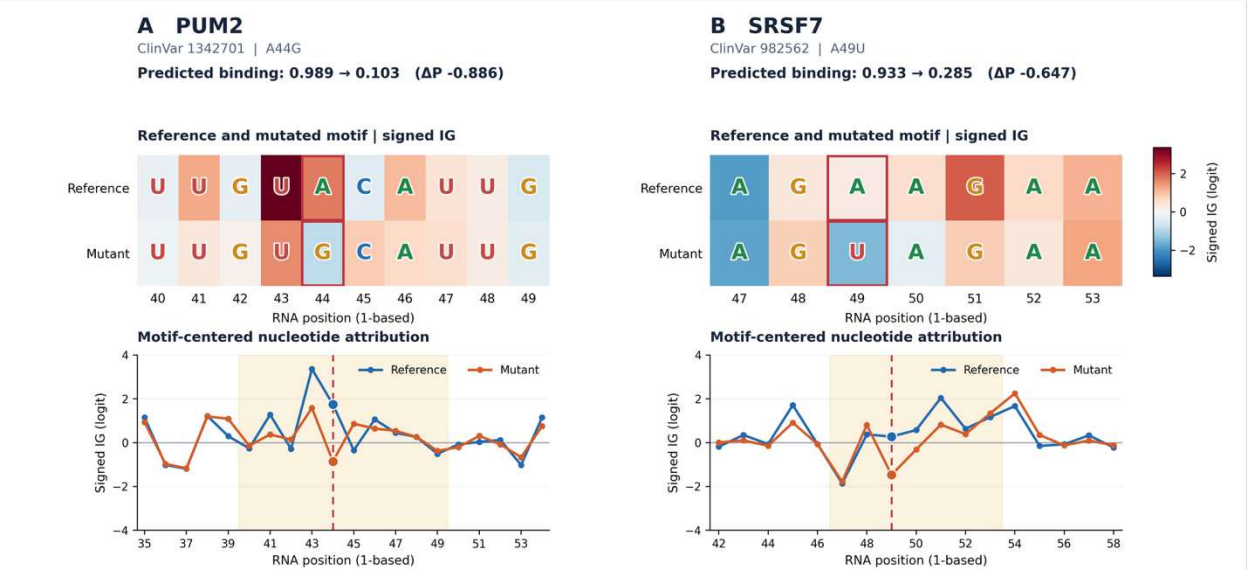


Fig. 8. Predicted effects of clinically annotated RNA variants on RBP binding. Reference and mutant RNA sequences were evaluated using GraPhiT-RBP to assess changes in predicted binding probability and nucleotide-level attribution. **(A)** PUM2-associated A44G substitution (ClinVar 1342701), resulting in a decrease in predicted binding probability from 0.989 to 0.103. **(B)** SRSF7-associated A49U substitution (ClinVar 982562), resulting in a decrease from 0.933 to 0.285. Upper heatmaps show signed Integrated Gradients (IG) attributions for nucleotides surrounding each substitution, with the mutated position highlighted. Lower plots compare reference and mutant nucleotide-level attribution profiles, with the substitution indicated by a dashed vertical line and the highlighted region showing the local sequence context. Attribution values represent model-derived feature contributions, not experimentally measured binding changes.

9. Residue-level cross-attention highlights RNA-binding domains and experimentally observed RNA-contact interfaces

To further investigate the biological interpretability of GraPhiT-RBP, we examined whether its residue-level cross-attention patterns correspond to protein regions known to participate in RNA recognition. Unlike approaches that compress protein representations into a single global vector before integrating them with RNA features, GraPhiT-RBP preserves individual amino acid representations throughout the Molecular Interaction Module. This enables the model to assign attention weights to individual protein residues in the context of a candidate RNA sequence, providing an opportunity to investigate how learned interaction information is distributed across different regions of an RBP.

We selected three representative RBPs with distinct RNA-binding domain architectures: TARDBP, PUM2 and PTBP1 (Fig. 9A–C). For each protein, we analyzed 200 positively predicted RNA sequences and extracted residue-level attention weights from the eight-head cross-attention module. Because RNA nucleotides serve as queries and protein residues provide keys and values, the attention weights describe how strongly individual RNA positions attend to amino acid residues. We averaged these weights across RNA positions and then across the 200 sequences to obtain a representative residue-level attention profile for each protein. We focused on attention head 8, which exhibited enrichment over annotated RNA-binding regions in the selected proteins. Protein structures predicted by AlphaFold were used to visualize the locations of these domains.

To quantify this correspondence, we calculated domain-attention enrichment as the mean attention assigned to residues within annotated RNA-binding domains relative to the

mean attention assigned to residues outside those domains. An enrichment value greater than 1 indicates that domain residues receive higher average attention than non-domain residues. Head 8 showed enrichment across all three representative proteins, with values of 1.752 for TARDBP, 2.761 for PUM2, and 1.300 for PTBP1. These correspond to approximately 1.75-fold, 2.76-fold, and 1.30-fold higher mean attention within annotated RNA-binding domains, respectively. PUM2 exhibited the strongest enrichment, consistent with prominent attention associated with its Pumilio homology domain (PUM-HD). Enrichment was also observed for the RNA recognition motifs of TARDBP and PTBP1 (Fig. 9A–C).

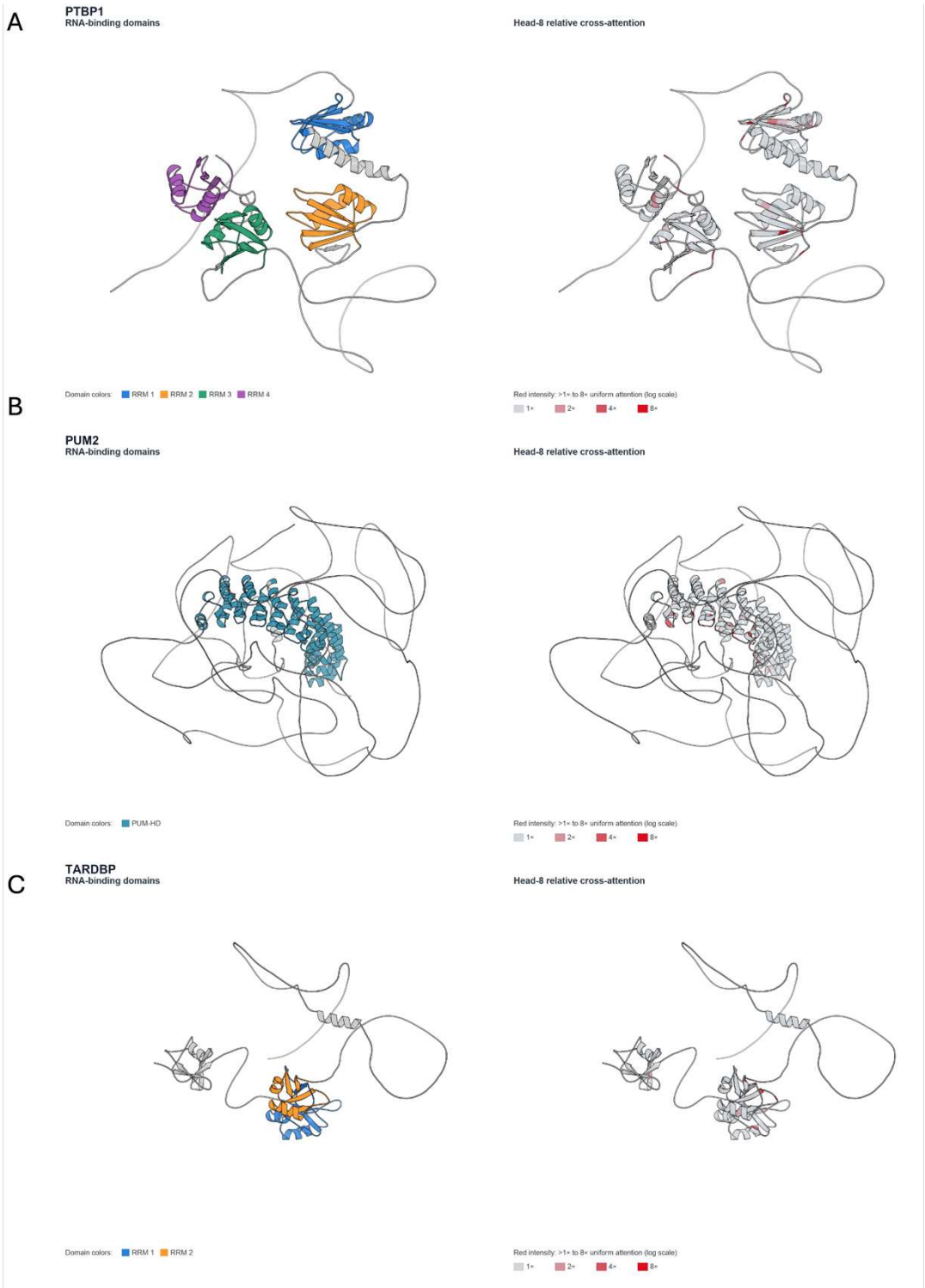


Fig. 9. Spatial alignment of learned cross-attention with annotated RNA-binding domains in representative RBPs. AlphaFold-predicted structures for (A) PTBP1, (B) PUM2, and (C) TARDBP are presented in a tandem layout to enable direct structural comparison. For each RBP, the left structure (Domain Map) highlights established RNA-binding domains using distinct categorical colors (PTBP1 RRM1–RRM4, PUM2 PUM-HD

and TARDBP RRM1/RRM2) with unannotated regions in gray. The right structure (Attention Map) displays the identical 3D orientation colored strictly by relative cross-attention. Residue-level attention profiles were extracted from attention head 8 of the GraPhiT-RBP Molecular Interaction Module—selected for demonstrating the highest overall enrichment for canonical binding motifs across all heads—and averaged across RNA nucleotide queries and 200 positively predicted RNA sequences. Increasing red intensity denotes higher attention relative to a uniform baseline (1× baseline; red saturation at 8×). Intrinsically disordered and low-confidence loop regions have been visually de-emphasized. Note that attention enrichment represents learned model weighting and should not be interpreted as direct evidence of physical RNA contact.

To further investigate whether highly attended residues correspond to experimentally observed RNA-binding interfaces, we mapped Head-8 attention onto the experimentally determined RNA-bound structures of TARDBP (PDB: 4BS2) and PUM2 (PDB: 3Q0Q) (Fig. 10A,B). Protein residues were classified as RNA contacts when any heavy atom was within 5 Å of an RNA heavy atom. We then selected the top 10% highest-attention residues within each experimentally resolved protein structure and examined their overlap with these RNA-contact sites. For TARDBP, 9 of the 17 highest-attention residues (52.9%) corresponded to experimentally observed RNA contacts, representing a 2.02-fold enrichment over the background contact frequency. Similarly, 18 of the 35 highest-attention residues (51.4%) in PUM2 contacted RNA, corresponding to a 4.08-fold enrichment. The structural visualizations further revealed spatial overlap between highly attended residues and experimentally defined RNA-binding interfaces (Fig. 10A,B).

Together, these findings indicate that GraPhiT-RBP's residue-level cross-attention preferentially highlights established RNA-binding domains and, in the examined experimental complexes, residues associated with RNA-contact surfaces. Although attention weights do not directly establish physical contacts or causal contributions to

binding, their correspondence with experimentally characterized interfaces provides additional evidence that the model captures biologically relevant protein regions during RNA–protein interaction learning.

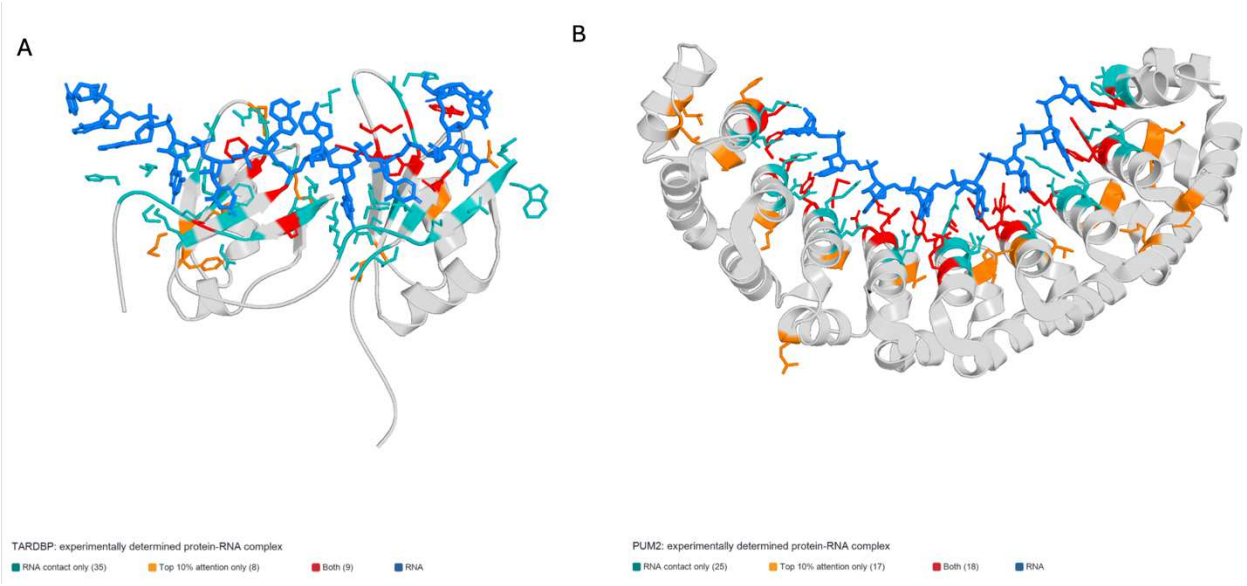


Fig. 10. Spatial correspondence between learned cross-attention and experimentally observed RNA–protein contacts. Experimentally determined RNA-bound structures of (A) TARDBP (PDB: 4BS2) and (B) PUM2 (PDB: 3Q0Q), showing protein residues classified according to their correspondence between Head-8 cross-attention and RNA-contact sites. RNA molecules are shown in blue, and the remaining protein structure in light gray. Teal indicates RNA-contact residues not included among the top 10% highest-attention residues, orange indicates highly attended residues not contacting RNA, and red indicates residues belonging to both groups. RNA contacts were defined using a minimum heavy-atom distance of 5 Å. Among the top 10% highest-attention residues, 9 of 17 TARDBP residues (52.9%) and 18 of 35 PUM2 residues (51.4%) contacted RNA, corresponding to 2.02-fold and 4.08-fold enrichment over their respective background contact frequencies. Attention scores were derived from 200 RNA sequences per protein. The overlap reflects spatial correspondence between learned attention and experimental RNA-binding interfaces rather than direct prediction of residue–RNA contacts.

Discussion

The growing availability of transcriptome-wide RNA-binding datasets has created opportunities to develop computational models that learn recognition patterns across multiple RNA-binding proteins (RBPs). However, learning these relationships remains challenging because different RBPs recognize distinct combinations of RNA sequence and structural features. Recent protein-aware approaches, particularly PaRPI, have advanced this field by allowing information to be shared across RBPs within a unified model. Building on this important development, we introduced GraPhiT-RBP to investigate whether a more compact architecture that directly models nucleotide–residue interactions could improve predictive performance and molecular interpretability. Our results demonstrate that prioritizing RNA–protein interaction learning, rather than relying on extensive independent RNA feature processing, provides an effective approach to protein-aware binding prediction.

A central contribution of GraPhiT-RBP is its ability to achieve stronger predictive performance with substantially reduced architectural complexity. Across 261 RBP datasets spanning six human cell lines, GraPhiT-RBP achieved the highest AUC on 206 datasets and outperformed PaRPI on 220 datasets, while reducing the number of trainable parameters by approximately 69%. Importantly, performance was evaluated separately for each RBP dataset using its corresponding test examples, and the resulting scores were subsequently summarized across datasets. Thus, the reported mean AUC reflects performance across individual protein-specific prediction tasks rather than being dominated by the largest datasets or calculated solely from predictions pooled across all

RBPs. This evaluation is particularly relevant for protein-aware models, since a strong aggregate prediction score does not necessarily indicate consistent performance across proteins with different RNA-recognition preferences. The broad improvements observed across individual datasets therefore support the ability of GraPhiT-RBP to learn useful recognition features for a diverse collection of RBPs.

One particularly interesting finding is that these improvements were obtained without requiring pretrained RNA language-model embeddings as an active model input. Whereas approaches such as HDRNet and PaRPI incorporate contextual RNA representations derived from BERT-based models, GraPhiT-RBP uses simple one-hot nucleotide identities together with experimentally informed icSHAPE reactivity and probabilistic RNA secondary-structure connectivity. Graph Transformer layers allow nucleotide information to propagate through both backbone and predicted base-pairing connections, providing structural context without requiring a separate pretrained RNA sequence encoder. The resulting RNA representation is then conditioned on individual protein residues through cross-attention before multiscale feature integration. The strong performance of this comparatively simple representation suggests that the effective integration of complementary molecular information can be more important than increasing the complexity of individual feature encoders. Rather than requiring every type of available RNA representation, the model demonstrates that a carefully organized combination of nucleotide identity, structural context, and protein-dependent interaction learning can provide highly informative features for binding prediction.

The architecture also highlights the importance of when molecular information is integrated. In PaRPI, RNA and protein representations are summarized independently before their interaction, whereas GraPhiT-RBP preserves individual nucleotide and amino acid representations during cross-attention. This allows the RNA representation to be conditioned on the particular protein being evaluated before sequence-level information is compressed. The ablation experiments support the importance of this design: removing protein embeddings, replacing cross-attention with additive fusion, or removing multiscale integration reduced predictive performance. The representation visualizations further illustrated how incorporating protein information reorganized RNA features according to binding status and protein identity. Together, these observations suggest that explicitly modeling the relationships between RNA positions and protein residues provides an effective alternative to increasingly elaborate RNA-only processing pipelines.

An important objective of protein-aware binding prediction is to extend learned recognition features beyond the specific RNA sequences and proteins encountered during training. GraPhiT-RBP demonstrated encouraging results in both settings. Under controlled RNA sequence similarity, it retained higher predictive performance than PaRPI across most evaluated RBP datasets. More importantly, in experiments where 22 proteins were completely excluded from training in each of K562 and HepG2, GraPhiT-RBP achieved higher AUC than PaRPI in 29 of 44 protein–cell-line evaluations. These results indicate that the model can use the residue-level representations of previously unseen proteins to generate binding predictions without requiring binding examples from those proteins during training. Although the magnitude of improvement was smaller than in the standard benchmark, the results are especially meaningful because predicting interactions for

previously uncharacterized RBPs represents one of the principal motivations for shared protein-aware modeling.

The transferability analyses provide further evidence that GraPhiT-RBP learns features that can be useful across different molecular and cellular contexts. Cross-cell evaluations between K562 and HepG2 demonstrated that models trained in one cell line could maintain predictive performance when applied to shared RBPs in the other, with performance broadly comparable to PaRPI. Analysis across MMseqs2-derived protein similarity groups further showed that the advantage of GraPhiT-RBP was not restricted to highly similar proteins. Notably, its relative performance improvements over PaRPI and HDRNet were greatest among the low-similarity protein group, supporting the value of the model's interaction-centered representation across RBPs with different sequence characteristics.

Another encouraging observation emerged when the number of proteins included during joint training was expanded. Increasing the training collection from 90 to 112 RBPs in K562 and from 62 to 84 RBPs in HepG2 improved predictive performance for selected proteins already represented in the original training sets. The illustrated examples also lacked qualifying MMseqs2 sequence-identity hits against the newly introduced proteins, suggesting that useful information can be gained from additional RBPs even when closely related protein sequences are not detected. These findings highlight the potential benefit of training protein-aware models on broader collections of RNA-binding proteins. Although the current benchmark includes 172 distinct RBPs, this remains a limited representation of the diversity of human RNA-binding proteins and their recognition

mechanisms. As larger experimentally characterized collections become available, incorporating more RBPs with different sequences, domain organizations, and binding preferences could enable models such as GraPhiT-RBP to learn an increasingly comprehensive representation of RNA–protein recognition.

Beyond predictive performance, GraPhiT-RBP provides several opportunities for investigating the molecular features associated with its predictions. Integrated Gradients analysis recovered recognizable RNA-binding sequence motifs, including the characteristic UGUA preference of PUM2 and GU-rich recognition patterns associated with TARDBP. Agreement with reference motifs from ATtRACT and mCross suggests that the model identifies nucleotide features consistent with established RNA-binding preferences. The analysis of clinically annotated single-nucleotide variants further illustrated how individual substitutions can substantially change predicted binding probabilities while altering nucleotide-level attribution in the surrounding sequence. These applications extend the model beyond binding classification by providing ways to identify potentially informative RNA sequence regions and prioritize variants for further biological investigation.

The preservation of protein residue information also enables an additional level of interpretation that is particularly relevant to RNA–protein recognition. Examination of cross-attention patterns revealed enrichment over annotated RNA-binding domains in representative proteins with different domain architectures, including TARDBP, PUM2, and PTBP1. Extending this analysis to experimentally determined RNA–protein complexes further revealed that highly attended residues preferentially localized to RNA-

contact interfaces. Among the top 10% highest-attention residues, 51.4% in PUM2 and 52.9% in TARDBP were within 5 Å of RNA, corresponding to 4.08-fold and 2.02-fold enrichment over their respective background contact frequencies. Notably, these patterns emerged despite the model being trained on only 172 distinct RBPs using binary RNA-binding labels, without explicit supervision from residue-level contacts or three-dimensional protein–RNA complexes. This suggests that the model may learn aspects of molecular recognition beyond those directly specified by its training objective. Although cross-attention weights do not establish physical contacts or causal contributions to binding, their correspondence with experimentally characterized interfaces provides a promising basis for identifying candidate protein residues involved in RNA recognition.

Several opportunities remain for extending the framework. The current model relies on contextual protein sequence embeddings without explicitly incorporating three-dimensional protein structure. Although these embeddings provide informative residue-level representations, RNA recognition may also depend on the spatial arrangement of amino acids, accessibility of binding surfaces, and the relative orientation of multiple RNA-binding domains. Incorporating protein structural connectivity and molecular-surface properties could therefore provide complementary information, particularly for proteins with complex recognition mechanisms or limited similarity to the training collection. Similarly, expanding the diversity of training RBPs and cellular conditions could strengthen the model's ability to generalize beyond currently characterized interactions. Experimental validation of predicted motifs, variant effects, and protein-residue contributions would further establish the biological relevance of the learned interaction features. Furthermore, expanding training to include a larger and more structurally diverse

collection of RBPs may improve the model's ability to capture transferable RNA-recognition features. Fine-tuning on experimentally resolved protein–RNA complexes with residue-level contact annotations could further refine these representations, potentially extending GraPhiT-RBP from interpretable binding prediction toward explicit identification of RNA-contacting protein residues.

Overall, GraPhiT-RBP demonstrates that protein-aware RNA-binding prediction can benefit from a more direct and efficient treatment of molecular interactions. By combining relatively simple RNA inputs with graph-based structural learning, residue-level protein embeddings, cross-attention, and multiscale integration, the framework improves predictive performance while retaining information useful for biological interpretation. Its performance across diverse RBP datasets, previously unseen proteins, and additional transferability evaluations illustrates the potential of learning shared RNA-recognition features across proteins. As larger and more diverse binding datasets become available, this interaction-centered approach provides a foundation for developing increasingly generalizable and biologically informative RNA–protein binding models.

Methods

Benchmark datasets and data partitioning

GraPhiT-RBP was evaluated using the established benchmark collection of 261 RNA-binding protein (RBP) binding datasets previously assembled for HDRNet [20] and subsequently used to evaluate PaRPI[21]. The benchmark comprises 172 human RBPs profiled across six cell lines, including K562, HepG2, HEK293, HEK293T, HeLa, and H9.

The collection contains 196 eCLIP datasets representing 111 RBPs from the ENCODE project[27] and 65 CLIP-seq datasets representing 61 RBPs from the POSTAR database[28]. For each RBP dataset, experimentally identified binding peaks were treated as high-confidence candidate binding regions, and the 5,000 peaks with the highest signal enrichment were retained as positive samples. Each region was standardized to a length of 101 nucleotides by symmetrically extending regions shorter than 101 nt or trimming longer regions from both ends. Negative samples consisted of 10,000 randomly sampled 101-nt transcriptomic regions, with regions overlapping known binding peaks of the corresponding RBP excluded. Positive and negative samples were assigned binary labels of 1 and 0, respectively.

To enable a direct and controlled comparison with previously reported benchmark results, GraPhiT-RBP preserved the same predefined 80:20 training-evaluation-testing split used in the PaRPI benchmark rather than generating new random data partitions. Thus, each RNA window was assigned to the same training or test partition used for the corresponding PaRPI evaluation. Following the cell-line-specific training strategy established by PaRPI, the training partitions of all RBP datasets belonging to the same cell line were combined to construct a single cell-line-specific training set. Six independent GraPhiT-RBP models were therefore trained, one for each cell line. The corresponding predefined test partitions were retained according to their original RBP dataset identities and were used to calculate performance separately for each of the 261 benchmark datasets. This fixed-split design ensured that GraPhiT-RBP and the previously reported benchmark methods were evaluated on matched test examples and prevented differences in random data partitioning from confounding model comparisons.

Additional controlled datasets were constructed to evaluate RNA- and protein-level generalization. For RNA generalization, we followed the same strategy used in previous studies[20, 21] and applied MMseqs2 to remove test RNA sequences sharing greater than 80% sequence identity with any training sequence. For protein generalization, dedicated holdout experiments were conducted independently in K562 and HepG2. In each cell line, 22 RBPs were completely excluded from training and retained exclusively for testing, while the remaining proteins—90 in K562 and 62 in HepG2—were used for model training. No binding samples from the 22 held-out proteins in either cell line were included during training, allowing us to evaluate the model's ability to generalize to previously unseen RBPs.

For comparative benchmarking, GraPhiT-RBP was evaluated against eight previously published RNA-binding prediction methods: PaRPI[21], HDRNet[20], PrismNet[19], PRIESSTESS[16], iDeep[14], GraphProt[15], DeepBind[13], and DMSK[17]. PaRPI was selected as the principal comparator because, like GraPhiT-RBP, it supports joint learning across multiple RBPs within a cell-line-specific model. During independent reproduction of PaRPI, we identified and corrected several implementation issues that adversely affected its predictive performance. These corrections produced improved results relative to the originally published scores. Rather than using the lower reported values, we adopted the corrected PaRPI results for our primary comparisons, ensuring that GraPhiT-RBP was evaluated against a stronger implementation of the most directly comparable method.

RNA and protein feature generation and precomputation

GraPhiT-RBP represents each RNA–protein pair using nucleotide-level sequence and structural information together with pretrained residue-level protein embeddings. To reduce computational overhead during training, RNA base-pairing probability matrices (BPPMs) and protein embeddings were generated in advance and stored for subsequent retrieval. The RNA graphs were then constructed dynamically during data loading using the precomputed structural information, without repeating RNA folding calculations.

For RNA, each sequence was standardized to 101 nucleotides, with the central 99 positions retained for graph construction. Each nucleotide was encoded using a four-dimensional one-hot representation corresponding to A, C, G, and U. Nucleotide-resolution *in vivo* icSHAPE reactivity values[18] were incorporated as an additional feature describing local RNA structural accessibility, producing a five-dimensional feature vector for each RNA graph node. Missing or invalid icSHAPE values were replaced with zero.

RNA secondary-structure information was generated using ViennaRNA/RNAfold[29]. For each 101-nt sequence, minimum-free-energy folding and partition-function calculations were performed to obtain a BPPM describing the probability of pairing between nucleotide positions. These matrices were computed in parallel before training, standardized to 101 × 101 dimensions, and stored in compressed pickle files for reuse. During data loading, each BPPM was retrieved and restricted to its central 99 × 99 positions. Base-pairing probabilities were then squared, and values below 0.04 were discarded. The RNA graph contained bidirectional edges connecting adjacent nucleotides along the backbone, together with additional edges corresponding to retained base-pairing interactions. Each edge was assigned three features: an indicator of sequential adjacency, the squared

base-pairing probability, and the sequence separation between the connected nucleotides normalized by the 99-position graph length. This graph representation captured both local backbone connectivity and probabilistic long-range secondary-structure interactions.

For proteins, pretrained ESM-3[30] was used to generate contextual embeddings for individual amino acid residues. Each residue was represented by a 1,536-dimensional embedding, preserving position-specific protein information without applying global sequence pooling. These embeddings were precomputed for each protein and stored in a protein-indexed pickle file. During data loading, the corresponding embedding matrix was retrieved using the protein identifier and retained at residue resolution for subsequent projection and RNA–protein cross-attention.

Overall, GraPhiT-RBP separates computationally expensive molecular feature generation from model training. RNA folding calculations and protein embedding generation are performed once during preprocessing, while nucleotide features and graph connectivity are assembled during data loading from the stored information. This approach avoids redundant RNA folding and protein language-model inference across training epochs while preserving the nucleotide- and residue-level representations required for interaction learning.

GraPhiT-RBP architecture and multimodal interaction learning

GraPhiT-RBP predicts RNA–protein binding through three sequential components (Fig. 1): the Molecular Representation Module, which encodes RNA sequence and structural

information using a graph Transformer and projects residue-level ESM-3 protein embeddings; the Molecular Interaction Module, which employs cross-attention between RNA nucleotide representations and protein residues; and the Interaction-Aware Multiscale Integration module, which integrates the resulting protein-conditioned RNA features for binary binding classification.

Molecular Representation Module

The RNA graph is defined as $G = (V, E, X, F)$, where V contains $N = 99$ nucleotide nodes corresponding to the central positions of each 101-nt sequence, E contains directed backbone and predicted base-pairing connections, $X \in \mathbb{R}^{99\times5}$ contains four-dimensional nucleotide one-hot identity and one icSHAPE reactivity value per node, and F denotes the edge-attribute matrix. The 101 × 101 base-pairing probability matrix (BPPM), precomputed using ViennaRNA, is cropped to the central 99 positions. For nucleotides i and j, the retained squared base-pairing probability is

$$s_{ij} = \begin{cases} p_{ij}^2, & p_{ij}^2 \geq 0.04, \\ 0, & \text{otherwise.} \end{cases}$$

Here, p_{ij} is the partition-function-derived pairing probability. The graph includes an edge when $|i - j| = 1$ or $s_{ij} > 0$. Each edge carries three features:

$$\mathbf{e}_{ij} = \left[\mathbb{1}(|i - j| = 1),\ s_{ij},\ \frac{|i - j|}{99}\right].$$

The three terms indicate sequence adjacency, retained pairing strength, and normalized sequence separation. Each five-dimensional node feature is first projected to 128

dimensions. Four edge-aware graph Transformer layers, each using four attention heads, then update the nucleotide features as

$$H^{(\ell+1)} = \mathrm{LN}\left(H^{(\ell)} + \mathrm{Dropout}\left[\mathrm{GELU}\left(\mathrm{TransformerConv}\left(H^{(\ell)}, E, F\right)\right)\right]\right).$$

Here, $H^{(\ell)} \in \mathbb{R}^{99\times128}$ denotes the nucleotide representations at layer ℓ, E specifies graph connectivity, and F contains the three-dimensional edge attributes. LN denotes layer normalization and GELU denotes the Gaussian error linear unit activation. Within each TransformerConv layer, node representations are projected into query, key, and value vectors for each attention head k, while the edge attributes are projected into the same feature space:

$$q_i^{(k)} = W_Q^{(k)} h_i^{(\ell)}, \qquad k_j^{(k)} = W_K^{(k)} h_j^{(\ell)},$$

$$v_j^{(k)} = W_V^{(k)} h_j^{(\ell)}, \qquad r_{ij}^{(k)} = W_E^{(k)} f_{ij}.$$

The attention coefficient assigned to neighboring nucleotide j when updating nucleotide i is calculated as

$$\alpha_{ij}^{(k)} = \mathrm{softmax}_{j\in\mathcal{N}(i)}\left(\frac{(q_i^{(k)})^{\top}\left(k_j^{(k)} + r_{ij}^{(k)}\right)}{\sqrt{d_h}}\right),$$

where $\mathcal{N}(i)$ denotes the graph-connected neighbors of nucleotide i and $d_h = 32$ is the dimensionality of each attention head. The resulting attention-weighted messages are aggregated as

$$m_i^{(k)} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)} \left(v_j^{(k)} + r_{ij}^{(k)} \right).$$

Thus, the edge attributes contribute to both attention-weight calculation and message aggregation by modifying the key and value representations, respectively. The outputs of the four attention heads are concatenated and combined with a learned transformation of the central node representation within TransformerConv. GELU activation and dropout are subsequently applied, followed by a residual connection and layer normalization. The resulting RNA graph representation is projected to 256 dimensions for interaction learning.

For a protein of length L, the pretrained ESM-3 residue embedding matrix $E_p \in \mathbb{R}^{L \times 1536}$ is projected into a 256-dimensional feature space as

$$P = E_p W_p + b_p, \qquad P \in \mathbb{R}^{L \times 256}.$$

Here, $W_p \in \mathbb{R}^{1536 \times 256}$ and $b_p \in \mathbb{R}^{256}$ are learnable parameters. This projection preserves a separate representation for each amino acid residue, without early global pooling or explicit protein structural graph construction, allowing residue-level features to participate directly in RNA–protein interaction learning.

Molecular Interaction Module

The Molecular Interaction Module applies eight-head cross-attention with RNA nucleotide representations as queries and protein residue representations as keys and values. Let $R \in \mathbb{R}^{99\times256}$ denote the graph-derived RNA representation and $P \in \mathbb{R}^{L\times256}$ the projected protein representation. For each attention head h, the normalized inputs are linearly transformed as

$$Q_h = \mathrm{LN}(R)W_h^Q, \qquad K_h = \mathrm{LN}(P)W_h^K, \qquad V_h = \mathrm{LN}(P)W_h^V.$$

Here, LN denotes layer normalization. Each head has dimension $d_k = 256/8 = 32$. The attention weight matrix and the corresponding head output are calculated as

$$A_h = \mathrm{softmax}\left(\frac{Q_h K_h^{\mathsf{T}}}{\sqrt{d_k}} + M\right), \qquad O_h = A_h V_h.$$

The softmax is applied across protein residues for each RNA nucleotide query. The additive mask M assigns negative infinity to padded protein positions and zero to valid residues, excluding padding from attention. The eight head outputs are concatenated and linearly projected to form the cross-attention update:

$$O = \mathrm{Concat}(O_1, \ldots, O_8)W^O.$$

A residual cross-attention update, followed by a position-wise feed-forward network (FFN), produces the protein-conditioned RNA representation:

$$R' = R + \mathit{Dropout}(O), \qquad R_{int} = R' + \mathit{Dropout}(\mathit{FFN}(\mathit{LN}(R'))).$$

Here, R denotes the original graph-derived RNA representation entering the Molecular Interaction Module, O is the projected cross-attention output incorporating protein information, and R′ represents the RNA features after the first residual update. The FFN comprises two linear layers with a GELU activation and dropout between them, and its output is added to R′ through a second residual connection. The module updates only RNA nucleotide representations; it does not perform reciprocal attention with protein residues as queries. Consequently, Rint ∈ $\mathbb{R}$99×256 retains one protein-conditioned feature vector per nucleotide before subsequent multiscale integration and global compression.

Interaction-Aware Multiscale Integration (IMI)

The IMI consists of three DPRBP-inspired pyramid stages. Let $X^{(0)} = R_{\mathrm{int}}^{\top} \in \mathbb{R}^{256\times 99}$ denote the protein-conditioned RNA features entering the module. At each stage s, the input undergoes one-dimensional max pooling (kernel size 3, stride 2), after one right-side padding position is assigned negative infinity. Two successive one-dimensional convolution–batch normalization–GELU operations refine the pooled features, and their output is combined with the pooled input through a residual connection:

$$\tilde{X}^{(s)} = \mathrm{MaxPool}_{3,2}\left(\mathrm{Pad}_{\mathrm{right}}\left(X^{(s-1)}, -\infty\right)\right),$$

$$X^{(s)} = \tilde{X}^{(s)} + \mathrm{Dropout}\left(\mathrm{ConvBlock}_{k_s}^{(2)}\left(\mathrm{ConvBlock}_{k_s}^{(1)}\left(\tilde{X}^{(s)}\right)\right)\right).$$

Here, $s \in \{1,2,3\}$ indexes the pyramid stages, $\tilde{X}^{(s)}$ denotes the pooled representation, and each ConvBlock consists of a one-dimensional convolution, batch normalization, and

GELU activation. The convolutional branch uses dropout (0.10). The stages use convolutional kernel sizes $k_s = 3,5,$ and 7, respectively, reducing RNA length from 99 to 49, 24, and 12 positions while retaining 256 channels.

Adaptive global max pooling across the remaining sequence positions produces a 256-dimensional interaction summary:

$$z = \mathrm{AdaptiveMaxPool}\big(X^{(3)}\big), \qquad z \in \mathbb{R}^{256}.$$

Finally, the summary is layer-normalized and passed through a 128-unit hidden layer with GELU activation and dropout (0.30), followed by a linear projection to a scalar binding logit. Sigmoid activation converts this logit into the predicted binding probability:

$$h = \mathrm{Dropout}_{0.30}\big(\mathrm{GELU}(W_1 \mathrm{LN}(z) + b_1)\big),$$

$$a = W_2 h + b_2, \qquad \hat{p} = \sigma(a).$$

Here, W_1 and b_1 parameterize the 256-to-128 hidden projection, W_2 and b_2 produce the scalar logit a, and σ denotes the sigmoid function. During training, the logit a is passed directly to binary cross-entropy with logits (BCEWithLogitsLoss), which incorporates sigmoid internally; explicit sigmoid activation is used to obtain binding probabilities during evaluation. This organization allows GraPhiT-RBP to integrate graph-derived RNA features with residue-level protein information before global sequence compression. Training parameters and the binary cross-entropy objective are described separately in the model training subsection.

Model training and computational implementation

GraPhiT-RBP was implemented using PyTorch and PyTorch Geometric and trained independently for each of the six cell-line-specific datasets. RNA–protein binding prediction was formulated as a binary classification task and optimized using binary cross-entropy with logits (BCEWithLogitsLoss). To account for differences in the proportions of binding and non-binding samples, the positive-class weight was calculated as the ratio of negative to positive samples in the training set and constrained to the interval [0.5, 2.0]. Model parameters were optimized using AdamW with a maximum learning rate of 8×10^{-5} and weight decay of 0.10. A OneCycle learning-rate schedule was applied over a maximum of 50 epochs, with the learning rate increasing during the first 20% of optimization steps. Gradient norms were clipped to a maximum norm of 1.0 to improve training stability.

Training was performed using automatic mixed precision, with bfloat16 used on compatible GPUs and float16 otherwise. Multi-GPU training was implemented using PyTorch DistributedDataParallel (DDP) with the NCCL communication backend, assigning one process to each GPU and distributing training samples through distributed samplers. A batch size of 48 samples per GPU process was used, together with 16 data-loading workers per process, pinned memory, and persistent workers. Precomputed RNA base-pairing probability matrices and residue-level ESM-3 protein embeddings were loaded from stored feature files, while RNA graph connectivity and edge attributes were constructed dynamically during data loading. This strategy avoided repeated RNA folding calculations and protein language-model inference during training.

Following the evaluation protocol adopted from PaRPI, model performance was assessed after each epoch using the predefined evaluation partition. The checkpoint achieving the highest evaluation AUC was retained, with early stopping applied when AUC failed to improve for 20 consecutive epochs, subject to the maximum of 50 epochs. The classification threshold was selected by maximizing the F1 score on the same evaluation partition, and both the selected checkpoint and corresponding threshold were saved for subsequent analyses. Random seeds were initialized for Python, NumPy, and PyTorch, with a process-dependent seed offset for distributed training.

Performance evaluation and comparative benchmarking

Predictive performance was evaluated separately for each of the 261 RBP datasets across six cell lines. The area under the receiver operating characteristic curve (AUC) was used as the primary threshold-independent metric. Classification performance was additionally assessed using accuracy, F1 score, and Matthews correlation coefficient (MCC). The area under the precision–recall curve (AUPRC) and recall were also calculated during model evaluation. Classification thresholds were determined by maximizing the F1 score over candidate probability thresholds from 0.10 to 0.89 in increments of 0.01. These thresholds were used to convert predicted binding probabilities into binary classifications for the corresponding evaluation.

GraPhiT-RBP was compared with eight previously published RNA-binding-site prediction methods: PaRPI, HDRNet, PrismNet, PRIESSTESS, iDeep, GraphProt, DeepBind, and DMSK. The predefined benchmark sample partitions were preserved to maintain consistency with previous evaluations. PaRPI was selected as the principal protein-aware

comparator because it also supports shared learning across multiple RBPs within individual cell lines, enabling comparison of the ability to learn protein-dependent RNA-binding patterns.

For each metric, performance was summarized across the 261 RBP-specific datasets, and the number of datasets on which each method achieved the highest score was recorded. Pairwise differences between GraPhiT-RBP and competing methods were calculated separately for each dataset. Statistical significance was assessed using paired Wilcoxon signed-rank tests across matched RBP datasets. Cell-line-specific comparisons were also conducted for H9, HEK293, HEK293T, HeLa, HepG2, and K562 to examine the consistency of performance differences across distinct experimental datasets. Results were summarized using AUC, accuracy, F1 score, and MCC, with additional comparisons of protein-level and RNA-level generalization presented separately.

Cross-cell transfer, protein sequence similarity, and training diversity

To further assess the transferability of GraPhiT-RBP, we conducted cross-cell evaluation between K562 and HepG2 using 62 RBPs shared by the two cell lines. Models trained on RNA-binding datasets from one cell line were evaluated on the corresponding RBP datasets from the other, with both K562-to-HepG2 and HepG2-to-K562 transfer examined. This analysis assessed whether protein-dependent RNA-binding patterns learned in one cellular context could support predictions in another.

To investigate the relationship between predictive performance and protein sequence similarity, we followed the protein homology analysis established by PaRPI. Protein

sequences were grouped using MMseqs2-based sequence comparisons, and performance was analyzed across the same three similarity categories: low (less than 30%), medium (30–60%), and high (greater than 60%). The original PaRPI similarity-group assignments were retained to maintain consistency with the published benchmark, allowing GraPhiT-RBP performance to be compared with existing methods within the same protein similarity groups.

Finally, we investigated the effect of expanding protein diversity during joint training. For K562, the original training collection containing 90 RBPs was expanded by introducing 22 additional RBPs, producing a training collection of 112 proteins. Similarly, the HepG2 training collection was expanded from 62 to 84 RBPs. Separate models were trained using the original and expanded protein collections, and performance was compared on the same original target RBP datasets to determine whether additional protein coverage affected predictions for previously represented proteins.

To examine whether observed improvements could be attributed simply to the introduction of closely related proteins, the sequences of the original target RBPs were compared with those of the 22 newly added RBPs using MMseqs2. Ten representative RBPs showing improved predictive performance and no qualifying sequence-identity hits against the added proteins were selected for visualization, with five examples from each cell line. This analysis provided illustrative cases in which performance improvements accompanied broader protein training coverage without detected sequence similarity to the newly introduced RBPs under the applied matching criteria.

Model ablation studies

Ablation experiments were performed using 17 HeLa RBP datasets to assess the contribution of individual architectural components. Four modified configurations were compared with the complete GraPhiT-RBP model: removal of the RNA graph representation, removal of pretrained protein embeddings, replacement of RNA–protein cross-attention with additive feature fusion, and removal of the DPRBP-inspired multiscale integration module. Each configuration was evaluated using the same RBP benchmark datasets, and performance differences were summarized using protein-level mean AUC, accuracy, and F1 score.

RNA-binding motif discovery using Integrated Gradients

To assess whether GraPhiT-RBP identifies sequence features associated with RBP recognition, we used Integrated Gradients (IG) to assign nucleotide-level importance scores to the model's predicted binding logits. For each RNA graph, the four one-hot identity features (A, C, G and U) were interpolated between a uniform reference value of 0.25 per channel and the observed one-hot vector. The measured icSHAPE features, RNA edges and their attributes, and protein embeddings remained fixed during attribution. IG was approximated using 32 midpoint integration steps, as follows:

$$\mathrm{IG}_k(x) \approx (x_k - {x'}_k)\frac{1}{32}\sum_{m=1}^{32}\frac{\partial f\left(x^{(m)}\right)}{\partial x_k^{(m)}}, \qquad x^{(m)} = x' + \frac{m-\frac{1}{2}}{32}(x - x').$$

Here, x is the observed RNA identity input, x' is the uniform baseline, f is the model's binding logit, k denotes a nucleotide-identity input feature, and $x^{(m)}$ is the interpolated RNA input at midpoint m. The four channel attributions were summed at each of the 99

modeled RNA positions to obtain signed positional attribution scores. To align these scores with the original 101-nucleotide RNA window, each of the 99 values was assigned to three consecutive original positions, and the values at each original position were averaged over their contributing modeled positions. The result was a signed, 101-position attribution profile.

Motif analysis was performed separately for the K562 RBP datasets. For each RBP, we selected up to 1,000 experimentally labeled binding sequences with predicted binding probabilities of at least 0.80, ranked by decreasing confidence, and up to 2,000 experimentally labeled non-binding sequences with predicted probabilities of at most 0.20, ranked by increasing prediction score. IG was calculated only for selected positive examples. For each positive 101-nt sequence, the mean signed IG score of every overlapping 6-mer was calculated. A candidate 6-mer was retained if its mean score was positive, exceeded the mean 6-mer score for that same RNA, and was at or above the 80th percentile of that RNA's 6-mer scores. Candidates were considered in descending score order and expanded to 21-nt contexts, approximately centered on the high-attribution 6-mer; overlapping 21-nt contexts within the same RNA were excluded.

For each RBP, these 21-nt regions formed the positive sequence set for motif discovery. Negative regions were sampled from distinct, confidently classified non-binding 101-nt sequences, matching the distribution of 21-nt starting positions observed in the positive set. The number of negative contexts was limited by the available negative sequences. Positive and negative contexts were supplied to STREME[31] independently for each

protein in RNA mode, searching motif widths of 5–10 nucleotides with a p-value threshold of 0.05.

Discovered STREME motifs were compared separately against reference motifs from the mCross[24] and ATtRACT[23] databases using TOMTOM[25], with Pearson correlation as the similarity metric, a minimum motif overlap of four positions, and a q-value reporting threshold of 0.05. TOMTOM q-values were used to assess motif-match significance after multiple-testing adjustment. Representative comparisons for PUM2, TARDBP, SRSF7 and TRA2A are presented in the Results.

Analysis of single-nucleotide variant effects on predicted binding

To evaluate the sensitivity of GraPhiT-RBP to single-nucleotide substitutions, clinically annotated variants from ClinVar[26] were introduced into corresponding reference RNA contexts. Reference and mutant 101-nt sequences differed only at the substituted nucleotide. RNA structure inputs were recomputed for the altered sequences, and the same trained model was used to estimate binding probabilities for each sequence pair. The predicted effect was defined as

$$\Delta P = P_{\mathrm{mutant}} - P_{\mathrm{reference}}.$$

Thus, negative values indicate reduced predicted binding following the substitution, whereas positive values indicate increased predicted binding. For nucleotide-level interpretation, the reference and mutant sequences were analyzed using IG with the same uniform 0.25 identity baseline and 32 midpoint steps described above. The resulting signed attribution scores were mapped to the original 101-nt coordinates and compared

at the substituted site and neighboring positions. Representative examples for PUM2 and SRSF7 were examined to illustrate changes in both predicted binding probability and local attribution. These are computational predictions and do not establish experimentally measured effects on RNA–protein binding.

Residue-level cross-attention and RNA-binding domain analysis

To investigate protein-residue contributions to the model's interaction representation, we analyzed the RNA-query/protein-key–value cross-attention matrices for TARDBP, PUM2 and PTBP1. For each protein, 200 positively predicted RNA sequences were used. For attention head 8, weights were first averaged across the 99 RNA query positions of each sequence, then averaged across the 200 sequences, yielding one mean attention score for each valid protein residue:

$$\bar{a}_j^{(8)} = \frac{1}{200}\sum_{s=1}^{200}\left(\frac{1}{99}\sum_{i=1}^{99} A_{ij}^{(s,8)}\right).$$

Here, $A_{ij}^{(s,8)}$ is attention from RNA position i to residue j for RNA sequence s in head 8. Relative residue-level attention was calculated against the uniform-attention expectation, namely $1/L$ for a protein with L valid residues:

$$r_j = L\bar{a}_j^{(8)}.$$

The residue scores were mapped onto AlphaFold-predicted protein structures and rendered in PyMOL. Annotated RNA-binding domains were assigned distinct base colors. Residues with relative attention at or below 1× retained the corresponding domain color;

residues above 1× were progressively blended toward red on a logarithmic scale, reaching full red at 8×. This display preserves domain identity while showing local differences in attention within each domain and the remaining protein.

To quantify the relationship between attention and established RNA-binding regions, domain enrichment was calculated as the mean per-residue attention within annotated RNA-binding domains divided by the corresponding mean outside those domains:

$$E_{\text{domain}} = \frac{\frac{1}{|D|}\sum_{j\in D}\bar{a}_j^{(8)}}{\frac{1}{|O|}\sum_{j\in O}\bar{a}_j^{(8)}}.$$

Here, D and O are the sets of domain and non-domain residues, respectively. Values greater than one indicate relatively higher mean attention within annotated domains. RNA recognition motifs of TARDBP and PTBP1 and the Pumilio homology domain of PUM2. Attention enrichment was interpreted as a property of the model's learned weighting, not as direct evidence of physical RNA–protein contacts.

To further assess whether highly attended protein residues corresponded to experimentally observed RNA-binding interfaces, we analyzed RNA-bound structures of PUM2 (PDB: 3Q0Q) and TARDBP (PDB: 4BS2). Experimental protein sequences were aligned to the corresponding full-length sequences to map residue-level Head-8 attention scores onto the resolved structures. For each protein residue, the minimum distance between any protein heavy atom and any RNA heavy atom was calculated, and residues with distances below 5 Å were classified as RNA contacts. Residues with valid attention scores were ranked by their mean Head-8 attention, and the highest-ranking 10% within

each experimental structure were selected. Contact overlap was quantified as the proportion of these high-attention residues that were classified as RNA contacts, and enrichment was calculated relative to the background contact frequency among all mapped residues. Analyses were restricted to residues with valid sequence mappings and attention scores; for the NMR structure of TARDBP (4BS2), the first structural model was used.

Data availability

All datasets used in this study, including the benchmarking data, precomputed RNA structural features, pretrained protein embeddings, and trained model checkpoints, are publicly available on Zenodo at https://doi.org/10.5281/zenodo.24108081. Additional data underlying the benchmarking figures are provided in the Supplementary Tables.

Code availability

The GraPhiT-RBP source code and implementation are available at https://github.com/moss-lab/GraPhiT_RBP.

Acknowledgements

The authors acknowledge the insightful comments of colleagues at Iowa State University.

Author contributions

A.O.D. and W.N.M. conceived the study. A.O.D. designed and implemented the GraPhiT-RBP framework, precomputed the molecular features, performed the benchmarking, generalization, ablation, and interpretability analyses, prepared the figures, and wrote the

original draft of the manuscript. T.I. assisted with formal analysis and validation. W.N.M. supervised the project, acquired funding, and provided project administration and computational resources. All authors reviewed, edited, and approved the final manuscript.

Funding

National Institute of General Medical Sciences [R01GM133810] to W.N.M.

Supplementary information

All data underlying Figures 1–5 are provided in Supplementary Table S1.

References list

1. Van Nostrand, E. L.*, et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 https://doi.org/10.1038/s41586-020-2077-3 (2020).

2. Castello, A.*, et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 https://doi.org/10.1016/j.cell.2012.04.031 (2012).

3. Baltz, A. G.*, et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**, 674–690 https://doi.org/10.1016/j.molcel.2012.05.021 (2012).

4. Martinez, N. M.*, et al.* Alternative splicing networks regulated by signaling in human T cells. *Rna* **18**, 1029–1040 https://doi.org/10.1261/rna.032243.112 (2012).

5. Taliaferro, J. M., *et al.* RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Mol Cell* **64**, 294–306 https://doi.org/10.1016/j.molcel.2016.08.035 (2016).

6. Tuerk, C. & Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **249**, 505–510 https://doi.org/doi:10.1126/science.2200121 (1990).

7. Ray, D., *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* **27**, 667–670 https://doi.org/10.1038/nbt.1550 (2009).

8. Lambert, N., *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* **54**, 887–900 https://doi.org/10.1016/j.molcel.2014.04.016 (2014).

9. Licatalosi, D. D., *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 https://doi.org/10.1038/nature07488 (2008).

10. Hafner, M., *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 https://doi.org/10.1016/j.cell.2010.03.009 (2010).

11. Duffy, E. E., *et al.* Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Mol Cell* **59**, 858–866 https://doi.org/10.1016/j.molcel.2015.07.023 (2015).

12. Van Nostrand, E. L., *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508–514 https://doi.org/10.1038/nmeth.3810 (2016).

13. Alipanahi, B., Delong, A., Weirauch, M. T. & Frey, B. J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* **33**, 831–838 https://doi.org/10.1038/nbt.3300 (2015).

14. Pan, X. & Shen, H. B. RNA-protein binding motifs mining with a new hybrid deep learning based cross-domain knowledge integration approach. *BMC Bioinformatics* **18**, 136 https://doi.org/10.1186/s12859-017-1561-8 (2017).

15. Maticzka, D., Lange, S. J., Costa, F. & Backofen, R. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol* **15**, R17 https://doi.org/10.1186/gb-2014-15-1-r17 (2014).

16. Laverty, K. U.*, et al.* PRIESSTESS: interpretable, high-performing models of the sequence and structure preferences of RNA-binding proteins. *Nucleic Acids Research* **50**, e111–e111 https://doi.org/10.1093/nar/gkac694 (2022).

17. Li, H.*, et al.* circRNA-binding protein site prediction based on multi-view deep learning, subspace learning and multi-view classifier. *Briefings in Bioinformatics* **23**, https://doi.org/10.1093/bib/bbab394 (2022).

18. Spitale, R. C.*, et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 https://doi.org/10.1038/nature14263 (2015).

19. Xu, Y.*, et al.* PrismNet: predicting protein–RNA interaction using in vivo RNA structural information. *Nucleic Acids Research* **51**, W468–W477 https://doi.org/10.1093/nar/gkad353 (2023).

20. Zhu, H.*, et al.* Dynamic characterization and interpretation for protein-RNA interactions across diverse cellular conditions using HDRNet. *Nature Communications* **14**, 6824 https://doi.org/10.1038/s41467-023-42547-1 (2023).

21. Peng, L.*, et al.* PaRPI predicts RNA-Protein interactions from cross-protocol and cross-batch RNA-binding protein datasets. *Communications Biology* **8**, 1396 https://doi.org/10.1038/s42003-025-08807-0 (2025).

22. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In: *International conference on machine learning*). PMLR (2017).

23. Giudice, G., Sánchez-Cabo, F., Torroja, C. & Lara-Pezzi, E. ATtRACT-a database of RNA-binding proteins and associated motifs. *Database (Oxford)* **2016**, https://doi.org/10.1093/database/baw035 (2016).

24. Feng, H.*, et al.* Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Mol Cell* **74**, 1189–1204.e1186 https://doi.org/10.1016/j.molcel.2019.02.002 (2019).

25. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 https://doi.org/10.1186/gb-2007-8-2-r24 (2007).

26. Landrum, M. J.*, et al.* ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* **42**, https://doi.org/10.1093/nar/gkt1113 (2014).

27. ENCODE Project Consortium, a. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).

28. Zhao, W.*, et al.* POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic acids research* **50**, D287–D294 (2022).

29. Lorenz, R.*, et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 26 https://doi.org/10.1186/1748-7188-6-26 (2011).

30. Hayes, T.*, et al.* Simulating 500 million years of evolution with a language model. *Science* **387**, 850–858 https://doi.org/doi:10.1126/science.ads0018 (2025).

31. Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 https://doi.org/10.1093/bioinformatics/btab203 (2021).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData1.xlsx
- SupplementaryTableS1.xlsx