# RNA-Protein Interaction Prediction Using Hybrid Quantum-Classical Fusion

A K M Muhitul Islam* and Gopalakrishnan Subramanian*
*Department of Computer Science, The University of Iowa, Iowa City, IA 52242, USA*

Rahul Singh†
*Department of Computer Science, The University of Iowa, Iowa City, IA 52242, USA*
†*Corresponding Author, Email: rahul-singh@uiowa.edu*

**These authors contributed equally to this work.*

Predicting whether a non-coding RNA and a protein physically interact is central to understanding post-transcriptional gene regulation and for screening candidate RNA sequences against a target protein. Existing predictive methods combine protein and RNA embeddings via classical operations—dot products, concatenations, or attention mechanisms—that induce different forms of interaction between the representations. We describe an approach that encodes fixed-weight protein (ESM-2) and RNA (RNA-FM) embeddings in two separate qubit registers of a quantum circuit and then entangles the qubits between those registers. This construct provides a coupling mechanism for functional inference beyond homology. An eight-qubit (four qubits per register), eight-layer circuit produces 48 cross-register XX, YY and ZZ expectations for a nonlinear classifier and exceeds its classical counterpart based on several measures (AUROC 0.792 vs. 0.760) despite having no trainable parameters of its own. We benchmarked this variant against established RNA–protein interaction prediction methods under the same 5-fold, family-and-clan-disjoint protocol on RNAInterAct. With encoders matched for RPIembeddor and the dual- and cross-encoder comparisons, the proposed method achieved 0.792 AUROC, compared with RPIembeddor (0.768), a dual-encoder (0.775) and a cross-encoder (0.781). These experiments provide preliminary evidence that an entangling circuit is a viable approach for combining biological embeddings. Ablation studies support the usefulness of cross-register correlators and nonlinear readout, whereas adding qubits or circuit depth did not consistently improve performance.

*Keywords: RNA-Protein Interaction Prediction; Quantum Machine Learning; Hybrid Quantum-Classical Models; Quantum Circuits; Entanglement; Parameter-Matched Ablation; Foundation Model Embeddings.*

## 1. Introduction

RNA-protein interactions are the physical binding events through which RNA-binding proteins recognize specific RNA sequences and structures to regulate splicing, localization, stability, and translation [1]. Experimentally assaying the combinatorial space of candidate pairs is infeasible, so computational predictions can be particularly useful to prioritize candidates for follow-up experimental validation. The recent advent of biological foundation models has changed how sequence data is encoded into features for downstream prediction tasks. Protein language models such as ESM-2, trained on hundreds of millions of evolutionary sequences, capture contextual

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

biochemical and structural signal without requiring a resolved structure. RNA language models such as RNA-FM provide analogous representations for non-coding RNA. Such models produce a fixed-length embedding for an input sequence. For our problem, given a protein embedding and an RNA embedding, the task lies in determining how they could be combined so as to predict whether a pair of protein-RNA molecules interact.

### *1.1 Related Work*

A survey of experimental methods used for solving this problem can be found in [1]. Computational methods that have been proposed to address this problem can be classified into a few categories: (i) *Sequence- and structure-based*: classifiers score pairs typically using hand-crafted features, such as RPISeq's RNA 4-mer and protein 3-mer (7-letter-alphabet) composition vectors [2], or RPI-Pred's 132-feature vector combining RNA secondary-structure-element counts (stem, hairpin, loop, bulge) with protein secondary-structure composition [3]. (ii) *Network-guided deep learning*: incorporates known interaction-network topology [4]. (iii) *Dual-encoder methods*: score pairs by similarity in a shared embedding space, as in RPIPLM [5]. (iv) *Cross-encoder methods*: concatenate or cross-attend between embeddings, as in CORAL [6]. (v) *Fine-tuned transformer systems* attend over- or fine-tune embeddings directly, as in AptaTrans [7] and RNAInterAct [8].

The mechanisms employed to integrate information in each of the above methods have limitations. For instance, a dot product reduces the rich information in the embeddings to a single number (alignment score). Concatenation, on the other hand, does not impose any structure at all, leaving inference of the relationship entirely to downstream processing. Cross-attention forms explicit pairwise scores between the two sets of embeddings. But the coupling it applies is itself parameterized: these scores are computed through learned projections and depend not just on the data but also the parametrization of the learning machine. Quantum machine learning offers a different way to couple the two embeddings. Rather than combining embeddings with a predefined mapping (*e.g*. the dot product), an unconstrained operator (*e.g*. concatenation), or a learned operator (*e.g.* cross-attention [8]), each embedding can be used to set the rotation angles of a small quantum circuit—one register of qubits per molecule and then entangle the two registers together. Entanglement gives us a mechanism to co-involve the information present in the two embeddings within a single quantum state, guaranteeing joint dependence on both molecules by construction (Section 1.2 defines this mechanism concretely). Prior work on data re-uploading [9] and quantum-enhanced feature spaces [10, 11] motivates the use of entangling circuits to construct features for classification. The novelty of our work lies in applying this approach to protein–RNA interaction prediction, using a quantum circuit to fuse information from different biological embeddings.

### *1.2 Functional Inference Beyond Homology*

Computational inference of RNA–protein molecular function increasingly requires methods that move beyond homology-based similarity: many RNA-binding proteins recognize short or structurally mediated motifs, and many functional RNA elements exhibit convergent protein-binding behavior without clear sequence similarity. Consequently, using homology as the only explanatory mechanism is limiting for understanding the diversity of RNA–protein binding events.

Foundation models address part of this gap by embedding sequences into representations that capture contextual signals that may go beyond explicit homology. However, the fusion step by which two such embeddings are combined has limitations as discussed earlier.

Our method uses quantum entanglement as a principled, parameter-free alternative to address this problem: once two registers are entangled, any measurement of any single qubit cannot be attributed to one molecule alone. Thus, joint dependence is enforced in a manner that is different than what may be learned from the data using parametrized learning machines. A qubit is the quantum analogue of a classical bit, and a register is simply the small set of qubits containing a molecule's embedding (four qubits per molecule in our implementation; Section 2.2). Two registers become entangled when a gate such as the cross-register CNOT (Section 2.2) acts on a superposition input, producing a joint state that cannot be factored in terms of the product of single-register states; a consequence is that the measurement outcomes of the two registers become statistically correlated.

### *1.3 Formulation of the Prediction Problem*

Let $P$ denote a protein sequence and $R$ an RNA sequence, drawn from the spaces of valid protein and RNA sequences respectively. Let $y(P,R) \in \{0,1\}$ indicate whether the pair physically interacts, with $y = 1$ denoting a true interaction and $y = 0$ a non-interacting pair. Given a labeled dataset of $n$ protein-RNA pairs with known interaction status, the goal is to learn a scoring function $f$ that maps a protein-RNA pair to an estimate $\hat{y} = f(P,R) \in [0,1]$ of the interaction probability, thresholded to obtain a binary prediction. For example, given a specific RNA-binding protein $P$ and a pool of candidate RNA sequences $R_1, \ldots, R_k$, ranking each candidate by $\hat{y}(P, R_k)$ is precisely how such a model would be used to prioritize sequences for experimental follow-up. Different machine learning or statistical methods may be used to construct $f$ based on $P$ and $R$. We describe the construction of a small quantum circuit that forms the core of a quantum machine learning method to construct $f$. The quantum circuit consists of two registers, one seeded by the RNA embedding and one by the protein embedding, and entangled by cross-register CNOT gates, such that every measured qubit reflects information from both molecules.

## 2. Methods

The inputs to our method are a protein sequence and an RNA sequence. By the term "embedding" hereafter, we imply the fixed-length, real-valued vector that each fixed, pretrained encoder produces for the corresponding sequence (Section 2.1 details both encoders). Given two such embeddings, our research investigated two questions: (1) whether routing the fusion step through an entangling quantum feature map produced a better interaction predictor than a (same-sized) classical network and (2) how the aforementioned approach compared with established protein-RNA interaction prediction methods. Answering this requires a classical model matched to the quantum model in every respect but the fusion mechanism, which is a central part of this study.

### *2.1. Fixed Weight Foundation-Model Encoders*

Both molecules were encoded using fixed weight, pretrained foundation models and neither of the encoders was fine-tuned by us. Protein sequences were encoded with a 150-million-parameter ESM-

2 model [12], mean-pooled over the sequence to a 640-dimensional vector $h_P$. RNA sequences were encoded with RNA-FM [13], a transformer language model pretrained on 23.7 million non-coding RNA sequences, mean-pooled to a 640-dimensional vector $h_R$. Both encoders operated on their native alphabets, so no nucleotide remapping was required. Every unique sequence in the dataset was encoded once, and the resulting vector was cached to disk, so training and evaluation only ever read pre-computed embeddings. This ensured that the fusion mechanism was sole variable under study. Equation (1) summarizes both encoders.

$$h_P = MeanPool\big(ESM2(P)\big) \in \mathbb{R}^{640}, h_R = MeanPool\big(RNA - FM(R)\big) \in \mathbb{R}^{640} \quad (1)$$

In addition to ensuring that the fusion mechanism was the sole variable under study, two reasons drove our decision to not fine-tune the encoders. First, by not fine tuning, every unique sequence could be embedded once and cached, so the shared-encoder comparisons involved the same input representations. Fine-tuning, conversely, would have let the encoders adapt differently under each fusion mechanism. Second, by not fine-tuning the encoders, we were able to keep the parameter-matched ablations computationally tractable, since re-encoding the dataset for every configuration would have increased the training cost substantially.

### *2.2. Quantum-Classical Fusion Architecture*

The quantum-fusion head, which we hereafter call Qfusion, turned the pair $(h_R, h_P)$ into a binding-probability logit in the following four stages.

*(1) Projection*:- Each embedding was projected by a dedicated two-layer MLP projector into a vector of raw angle features, one feature per (wire, rotation axis, circuit layer) triple assigned to that molecule's register. With 4 qubits per register, 8 re-uploading layers, and 3 encoding axes (RX, RY, RZ), each projector output 4 × 8 × 3 = 96 features (Eq. 2). Each projector used 128 hidden units with GELU, input and output layer normalization, and dropout 0.1 before each linear layer. This configuration was selected, from a qubit-count, re-uploading-depth and readout ablation evaluated under the same 5-fold protocol, as the strongest-performing quantum variant; the full ablations are reported in Sections 3.5 and 3.6.

$$a = f_R(h_R) \in \mathbb{R}^{96}, p = f_P(h_P) \in \mathbb{R}^{96} \quad (2)$$

*(2) Angle scaling:*- The raw features $x = [a \| p] \in \mathbb{R}^{192}$ were mapped to bounded rotation angles by a trainable, elementwise affine-tanh transform. This kept every angle inside $(-\pi, \pi)$ and let the model learn how sharply each feature modulated its rotation (Eq. 3).

$$\theta = \pi \tanh(w \odot x + b) \quad (3)$$

$$u_{l,k,q} = \begin{cases} a_{l,k,q}, & 1 \le q \le 4 \\ p_{l,k,q-4}, & 5 \le q \le 8 \end{cases}$$

In words: every angle was driven by exactly one tower — the RNA projector supplied the angles on wires 1 to 4 and the protein projector supplied the angles on wires 5 to 8, at every layer *l* and on every rotation axis *k*.

*(3) Entangled circuit:*- The 192 angles drove an eight-qubit circuit, four wires for the RNA register and four for the protein register, ran for eight data re-uploading layers on a statevector simulator. Within each layer: (*i*) every wire received one RX, one RY, and one RZ rotation, each

driven by a different re-uploaded angle from $\theta$; (ii) an intra-register ring of CZ gates entangled the four RNA wires among themselves, and separately entangled the four protein wires among themselves; (iii) four cross-register CNOT gates coupled RNA wire *i* to protein wire *i, for i = 1...4.* (A statevector simulator computed the circuit's exact quantum state—a vector of $2^8 = 256$ complex amplitudes for our eight-qubit circuit—on ordinary classical hardware, rather than executing the circuit on physical quantum hardware.) Step (iii) was the fusion step. The cross-register CNOT gates coupled the RNA and protein registers enabling entanglement and allowing joint information to be accessed through cross-register measurements. All representational learning occurred in the projectors, in the angle-scaling parameters $w$ and $b$, and in the readout layer described next. The input-dependent circuit rotations changed through their learned encoding angles.

We numbered the eight wires $q = 1,\ldots,8$, with wires 1 to 4 forming the RNA register and wires 5 to 8 the protein register, Fig. 1 depicts the entangled circuit including the pairing of RNA wire *i* with protein wire *i*. $L = 8$ is the number of re-uploading layers and $K = 3$ is the number of encoding axes. Throughout, $\sigma 1 = X$, $\sigma 2 = Y$, and $\sigma 3 = Z$ denote the Pauli matrices, and $R\sigma(t) = \exp(-it\sigma/2)$ denotes the rotation of a single qubit by angle $t$ about axis σ. Equation (4) defines the full block $S_l$.

$$S_l(\theta) = \left[\bigotimes_{q=1}^{8} R_Z\left(\theta_{l,3,q}\right)\right]\left[\bigotimes_{q=1}^{8} R_Y\left(\theta_{l,2,q}\right)\right]\left[\bigotimes_{q=1}^{8} R_X\left(\theta_{l,1,q}\right)\right], \quad W = W_{\times}(W_R \otimes W_P) \tag{4}$$

Operators acted right to left, so each wire received its RX rotation first, then RY, then RZ. Within one axis the eight rotations acted on distinct wires and commuted. The entangler $W$ was built from an intra-register CZ ring on each register, $W_R$ over wires 1 to 4 and $W_p$ over wires 5 to 8, followed by the four cross-register gates $W\times$ = CNOT(1→5) CNOT(2→6) CNOT(3→7) CNOT(4→8). The CZ gates were diagonal and so commuted among themselves, and $W_R$ and $W_p$ acted on disjoint registers, so $W\times$ was the only term that coupled the two molecules.

$$|\psi(\theta)\rangle = W\, S_L(\theta)\, W\, S_{L-1}(\theta) \cdots W\, S_1(\theta)\, |0\rangle^{\otimes 8} \tag{5}$$

The circuit started from $|0\rangle\otimes 8$ and alternated an encoding block with the entangler for $L = 8$ layers.

(4) *Readout:-* After eight layers we calculated XX, YY and ZZ expectations for every cross-register pair, $i, j = 1...4$. Each axis contributed sixteen correlators, giving forty-eight values in [−1, 1] (Eq. 6), concatenated in XX, YY, ZZ order. A 48-to-32-to-1 classifier with GELU and dropout 0.1 after the hidden activation mapped this vector to a scalar logit; a sigmoid gave the interaction score (Eq. 7). No single-wire values entered this readout. In Eq. (6), $A \in \{X,Y,Z\}$; $D_{0.1}$ in Eq. (7) was dropout during training and the identity at inference. The pair observables exposed joint measurement features to the classifier; Section 3.6 compared ZZ-only and XX/YY/ZZ features with linear and nonlinear classifiers. (An expectation value was the average outcome obtained if the same measurement were repeated many times; "XX" denoted jointly measuring two qubits along the Pauli-X axis, and similarly for YY and ZZ—the three standard ways of reading out a qubit pair's correlated state.)

$$z_{ij}^{(A)} = \langle\psi(\theta)|A_i A_{4+j}|\psi(\theta)\rangle \tag{6}$$

$$\hat{y} = \sigma(w_2^T D_{0.1}[GELU(W_1 z + b_1)] + b_2) \tag{7}$$

The full circuit used 288 gates (64 RX, 64 RY, 64 RZ, 64 CZ, 32 CNOT) at a circuit depth of 64. The model's 193,793 trainable parameters lived entirely in the two projectors, the angle-scaling

vectors, and the readout network (1,601 parameters). Training used binary cross-entropy between the predicted probability $\hat{y}$ and the ground-truth label $y$ (Eq. 8).

$$L = -[ylog(\hat{y}) + (1-y)log(1-\hat{y})] \quad (8)$$

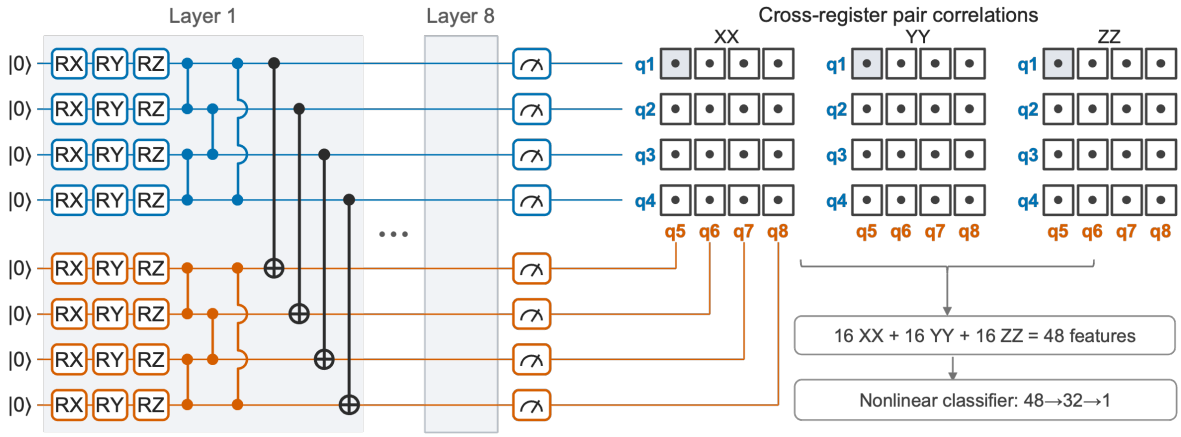



Fig. 1. Schematic representation of the Qfusion entangling circuit, showing one of the eight layers and the four-wire RNA and protein registers. RX, RY, and RZ gates encode the per-tower angle features; intra-register CZ gates form a ring within each register; cross-register CNOT gates couple each RNA control wire to the corresponding protein target wire. The final state yields 16 XX, 16 YY and 16 ZZ pair expectations. These 48 features enter a 48→32→1 nonlinear classifier with GELU and training-time dropout 0.1; a sigmoid converts the output logit to an interaction *score*.

### *2.3. Parameter-Matched Classical Control*

In the first experiment we compared the proposed architecture with a standard deep learning architecture (SDLA) consisting of two projector architectures, producing angle vectors $a$ and $p$ of the same dimension as above. These were passed through $tanh$ and scaled by $\pi$ without the learned affine transform in Eq. (3), but instead of entering a quantum circuit, the concatenated 192-dimensional vector was fed to a small classical fusion network. This was a 192-to-8-to-1 multilayer perceptron with a GELU nonlinearity, whose output was the logit. The width of this fusion layer set the control's parameter count; we fixed it by selecting the width whose total landed close to Qfusion's, which gave eight hidden units.

The SDLA had 193,361 trainable parameters, within 0.22% of Qfusion's 193,793. The models used the same pretrained embeddings and projector dimensions, with a classical fusion MLP in the control and a fixed entangling circuit followed by a nonlinear readout in Qfusion. This comparison evaluated the complete fusion architectures; matching parameter counts alone did not equate learning capacity or entanglement.

### *2.4. Baselines Retrained Under the 5-Fold Protocol*

We additionally trained and evaluated two classical fusion baselines with frozen encoders on the same embeddings, data, and 5-fold protocol. The dual-encoder ("Dual-encoder" in Table 1) projected each tower into a 256-dimensional embedding, L2-normalized both, and scored the pair by cosine similarity times a learned temperature, in the style of contrastive dual-tower models such

as RPIPLM [5]. The cross-encoder ("Cross-encoder") concatenated the two pooled embeddings and passed them through a three-layer MLP to a scalar logit.

Three further methods, whose performance had previously been known only from externally reported values, were retrained within this study on the same 5-folds. RPIembeddor [8], the reference model released with RNAInterAct, used the same fixed RNA-FM and ESM-2 150M embeddings as Qfusion, with cross-attention and convolutional fusion. The comparison in Table 1 was therefore encoder-matched with model-specific downstream architectures and optimization. AptaTrans [7] was evaluated in two variants. These were fixed weight pretrained encoders with only the prediction stage trained ("AptaTrans-fixed weight"), and full end-to-end fine-tuning ("AptaTrans-retrained"). All seven methods in Table 1 shared the same evaluation partitions, with their stated model-specific training settings.

### *2.5. Protocol for Training and Testing*

We evaluated all methods on the RNAInterAct benchmark [8], restricted to a family-and-clan-disjoint partition and extended to 5-fold cross-validation. RNA families and protein clans were each independently split into 5 groups and for fold $i$, the test set contained every pair whose RNA family and protein clan both fell in group $i$, the validation set used group $(i + 1) \bmod 5$, and the remaining three groups formed training. This guaranteed that every test RNA family and every test protein clan were entirely absent from that fold's training and validation data. Checkpoints were selected using each fold's disjoint validation group. This reduced annotated family and clan overlap but did not exclude all evolutionary signal in the embeddings. Across the 5 folds, training sets ranged from 34,758–35,500 pairs, and validation and test sets each ranged from 3,833–4,017 pairs, for 19,606 unique test pairs evaluated in total. Positive-pair prevalence varied more than fold size (23.3–59.9% across folds), since it depended on which families and clans happened to co-occur. There was zero sequence overlap between any fold's training, validation, and test sets for either molecule.

All trainable models were optimized with AdamW [14] and a cosine-annealing learning-rate schedule [15], using binary cross-entropy loss, and trained independently on each of the 5 folds. Qfusion and its four pair-readout variants used learning rate 0.001, weight decay 0.01, batch size 256, projector dropout 0.1, up to 25 epochs and early-stopping patience 10. The cosine schedule had a 100-epoch horizon. The retained baseline and circuit-size experiments used up to 100 epochs with patience 30, or up to 40 epochs with patience 12 for AptaTrans. We reported mean and standard deviation across folds for every method, rather than a single training seed.

Each fold provides its own family-and-clan-disjoint validation partition, so early stopping and checkpoint selection use that fold's validation AUROC. All four pair-readout configurations were completed and their checkpoints fixed before test evaluation. Architecture comparisons are exploratory because they use the same reported benchmark; one fold's validation group is another fold's test group.

For every model we reported AUROC and AUPRC across all thresholds, and at the 0.5 threshold F1, precision, recall, accuracy, balanced accuracy, and MCC, plus the Brier score for calibration. For Qfusion against its classical control, we ran a paired $t$-test, with a Wilcoxon signed-rank test as a distribution-free check, across the 5-fold-level values of each measure, and a McNemar test on paired 0.5-threshold decisions, both pooled across folds and separately within each fold, since

pooling can manufacture apparent significance from one unrepresentative fold. Full holistic measures required raw per-pair predictions, which we retained for the four models trained through the shared fusion pipeline: Qfusion, its classical control, and the dual- and cross-encoder baselines. Our circuit ran on a classical statevector simulator at a scale that is trivial to simulate using standard hardware, and we claimed no runtime or sample-complexity quantum advantage.

## 3. Results

### *3.1. Overall Predictive Performance*

Table 1 reports mean AUROC, AUPRC, and F1 (± SD across the 5 folds) for all seven methods, evaluated on the same family-and-clan-disjoint partitions. The primary comparison of this study was Qfusion, the eight-qubit, eight-layer configuration examined in the case studies in Sections 3.3–3.6, against established RNA-protein interaction methods. Under encoder-matched conditions Qfusion reached 0.792 mean AUROC, ahead of RPIembeddor (0.768) and the fixed weight dual-encoder (0.775) and numerically above the cross-encoder (0.781). A second, narrower comparison, against a classical fusion network matched to Qfusion's own encoders and parameter count, was examined in Section 3.3, where Qfusion exceeded that matched control (0.760 mean AUROC) on every measure evaluated. Section 3.3 examines the uncertainty of these differences. AptaTrans trailed the field under both training regimes tested, possibly reflecting differences from its narrower aptamer-protein setting (Section 4). The three stronger baselines were all designed for this same task, and all were evaluated here on the same fixed RNA-FM and ESM-2 150M embeddings that Qfusion used. Holding the encoders fixed across methods controlled their input representations, and under that constraint the entangling fusion head was competitive with the strongest of these baselines.

Table 1. Comparative evaluation across RNA-protein interaction prediction methods on the RNAInterAct family-and-clan-disjoint 5-fold cross-validation (5 folds; training 34,758–35,500 pairs, validation and test 3,833–4,017 pairs per fold). All seven rows were evaluated on these partitions with the model-specific training settings in Section 2.5; values are mean ± SD across the 5 folds. The proposed method (Qfusion) is listed last and shown in bold.

| Method | AUROC | AUPRC | F1 |
|---|---|---|---|
| RPIembeddor | 0.768 ± 0.036 | 0.656 ± 0.070 | 0.482 ± 0.079 |
| Cross-encoder | 0.781 ± 0.030 | 0.682 ± 0.115 | 0.603 ± 0.105 |
| Dual-encoder | 0.775 ± 0.023 | 0.676 ± 0.117 | 0.617 ± 0.087 |
| Classical control (matched) | 0.760 ± 0.041 | 0.650 ± 0.100 | 0.601 ± 0.037 |
| AptaTrans (retrained) | 0.608 ± 0.034 | 0.492 ± 0.118 | 0.537 ± 0.068 |
| AptaTrans (fixed weight) | 0.538 ± 0.032 | 0.446 ± 0.148 | 0.403 ± 0.178 |
| **Quantum-fusion (Qfusion)** | **0.792 ± 0.038** | **0.695 ± 0.090** | **0.642 ± 0.043** |

### *3.2. Holistic, Beyond-AUROC Assessment*

Table 2 and Figure 2 report the full evaluation suite for the four models trained through the shared fusion pipeline: Qfusion, its classical control, and the dual- and cross-encoder baselines, all of which retained raw per-pair predictions. Qfusion nominally led its classical control on every measure in Table 2, using an entangling circuit that contributed zero trainable parameters of its own, against a learned classical fusion network. At the 0.5 decision threshold Qfusion led its classical control on F1 (0.642 vs. 0.601), recall (0.633 vs. 0.575), MCC (0.404 vs. 0.353) and balanced accuracy (0.705 vs. 0.677). Qfusion also had higher mean F1 and MCC than the dual- and cross-encoder baselines, while the cross-encoder retained higher precision and a lower Brier score. Returning to the classical-control comparison that motivated Section 3.3, these margins were modest relative to the fold-to-fold standard deviations reported in Table 2 (±0.023 to ±0.119 depending on the measure). What made them worth reporting was that the direction of the effect was consistent across every one of the eight measures evaluated, rather than favorable on some and unfavorable on others. Section 3.3 examines matched-fold uncertainty.

### *3.3. Isolating the Quantum Contribution: Statistical Robustness*

We compared Qfusion with the classical control using the same encoders and projector dimensions and closely matched trainable parameter counts (193,793 vs. 193,361, within 0.22%). The architectural differences are specified in Section 2.3. This comparison complemented the established-method benchmarks in Section 3.1. We used paired tests across the five matched evaluation folds because their training sets overlapped, the resulting $p$-values were interpreted as nominal summaries.

Quantum fusion was ahead of its classical control on both ranking measures in all 5 folds. The AUROC differences ranged from 0.0016 to 0.0511. In its best-performing fold (fold 4), the AUROC value for Qfusion was 0.848 and comparisons with other methods used the matched-fold differences rather than their separate best-fold scores. A paired $t$-test gave a mean AUROC difference of +0.0316 ($p = 0.027$) and a mean AUPRC difference of +0.0452 ($p = 0.068$). The Wilcoxon signed-rank test gave $p = 0.063$ for each ranking measure. Mean F1 increased by 0.0410 (paired $p = 0.002$); recall, MCC and balanced accuracy also had nominal paired $p < 0.05$, whereas precision and Brier score did not.

At a fixed 0.5 decision threshold, the pooled McNemar test involved 19,606 pairs. Qfusion correctly classified 2,026 pairs that its classical control misclassified, versus 1,552 pairs where the classical control was correct and Qfusion was not (continuity-corrected statistic 62.53; nominal $p = 2.6 \times 10^{-15}$). Within each of the five folds, more discordant pairs favored Qfusion over the classical control; four folds had nominal $p < 0.05$, while fold 2 did not. Shared molecules and biological groups limited the independence assumption of the pooled test. It assessed thresholded errors, not AUROC or an entanglement-specific effect.

Overall, Qfusion's point estimates favored it on every reported measure against this control. The AUROC paired $t$-test and thresholded comparisons provided encouraging evidence, although the rank-based test and the dependence between folds limited the strength of inference. Against the cross-encoder, the mean AUROC difference was +0.0106 (paired $p = 0.414$), indicating that the entangling-circuit mechanism led to comparable or slightly improved performance.

### *3.4. Quantum Circuit Resource Profile*

Table 3 summarizes the circuit's resource footprint, for assessing the cost of the quantum approach. The circuit used 8 qubits, a width that could be easily simulated using classical hardware. This configuration was identified through the circuit and readout comparisons in Sections 3.5–3.6. We presented it as an architectural ablation, not a quantum-hardware demonstration. At eight qubits the simulated state had 256 amplitudes; the XX/YY/ZZ nonlinear runs took 16.6–22.3 minutes per fold on an RTX A6000-based machine.

Table 2. Full holistic measures at the 0.5 probability threshold, mean ± SD across the 5 family-and-clan-disjoint folds.

| Model | AUROC | AUPRC | F1 | Prec. | Recall | MCC | Bal. Acc. | Brier |
|---|---|---|---|---|---|---|---|---|
| **Quantum-fusion** | **0.792 ±0.038** | **0.695 ±0.090** | **0.642 ±0.043** | **0.665 ±0.095** | **0.633 ±0.071** | **0.404 ±0.054** | **0.705 ±0.036** | **0.220 ±0.045** |
| Classical control | 0.760 ±0.041 | 0.650 ±0.100 | 0.601 ±0.037 | 0.646 ±0.109 | 0.575 ±0.059 | 0.353 ±0.051 | 0.677 ±0.031 | 0.244 ±0.058 |
| Dual-encoder | 0.775 ±0.023 | 0.676 ±0.117 | 0.617 ±0.087 | 0.660 ±0.119 | 0.585 ±0.080 | 0.381 ±0.042 | 0.690 ±0.028 | 0.230 ±0.030 |
| Cross-encoder | 0.781 ±0.030 | 0.682 ±0.115 | 0.603 ±0.105 | 0.677 ±0.103 | 0.549 ±0.118 | 0.375 ±0.040 | 0.680 ±0.028 | 0.211 ±0.037 |

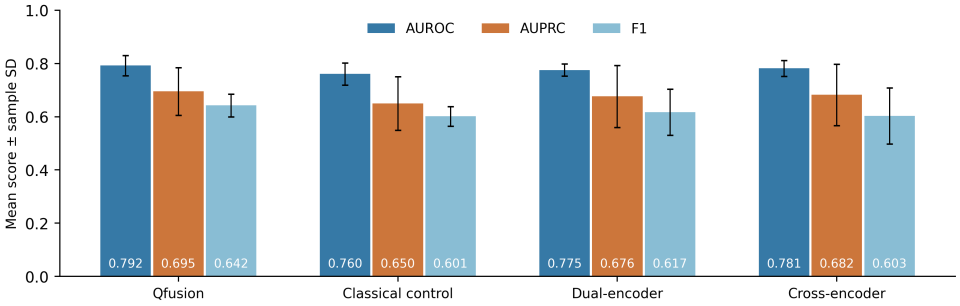



Fig. 2. AUROC, AUPRC, and F1 (mean ± SD, error bars) for the four models across the 5 family-and-clan-disjoint folds.

Table 3. Quantum circuit resource summary.

| Property | Value |
|---|---|
| Qubits | 8 (4 RNA register + 4 protein register) |
| Data re-uploading layers | 8 |
| Encoding axes per wire per layer | 3 (RX, RY, RZ) |
| Readout observables | 48 (16 XX + 16 YY + 16 ZZ cross-register expectations) |
| Total gates | 288 (64 RX, 64 RY, 64 RZ, 64 CZ, 32 CNOT) |
| Circuit depth | 64 |
| Simulator | PennyLane default.qubit, exact statevector simulation |
| Qfusion trainable parameters (total) | 193,793 |

| Property | Value |
|---|---|
| Classical-control trainable parameters (total) | 193,361 |
| Parameter match | within 0.22% |

### *3.5. Qubit-Count and Re-uploading-Depth Ablation*

Six circuit configurations were trained and evaluated under the identical 5-fold protocol, spanning 2 to 12 qubits and 6 to 256 re-uploading layers per register, all using the single-wire (xyz) readout only. With a fixed readout type, changing width and depth also changed projector dimensions and parameter counts. The results from this experiment are in Table 4. Under a single-wire readout the quantum head plateaued near AUROC 0.76 across a wide range of widths (Table 4). The deepest configurations had lower mean AUROC, as low as 0.707, despite using up to three times the default parameter count. Neither added width nor added depth lifted the single-wire readout past its plateau, which motivated the readout ablation in Section 3.6. None of the six configurations in Table 4 outperformed the classical control from Table 1 ($0.760 \pm 0.041$); every one of them measured single-wire observables only. Section 3.6 examines cross-register correlators and classifier form as alternative ways to improve the readout.

Table 4. Qubit-count / re-uploading-depth ablation for Qfusion, mean ± SD AUROC across the 5 family-and-clan-disjoint folds. "Params" was the total trainable parameter count of the full model (projectors, angle-scaling, and readout). All six configurations used the single-wire (xyz) readout only, without the cross-register correlators of Section 3.6.

| Configuration | Qubits | Layers | Params | AUROC |
|---|---|---|---|---|
| 6q, 8 layers | 6 (3+3) | 8 (sym.) | 185,827 | $0.759 \pm 0.041$ |
| 4q, 128 layers | 4 (2+2) | 128 (sym.) | 370,957 | $0.759 \pm 0.014$ |
| 12q, 6 layers | 12 (6+6) | 6 (sym.) | 195,421 | $0.757 \pm 0.051$ |
| 4q, 256 layers | 4 (2+2) | 256 (sym.) | 575,245 | $0.746 \pm 0.028$ |
| 2q, 160 layers | 2 (1+1) | 160 (sym.) | 294,343 | $0.738 \pm 0.040$ |
| 2q, 256 layers | 2 (1+1) | 256 (sym.) | 370,951 | $0.707 \pm 0.066$ |

### *3.6. Readout Ablation: Cross-Register Correlators*

In Table 5 we present results from changing the measurement readout within the circuit family. The twelve-qubit circuit measured 36 single-wire observables, twice as many as the six-qubit circuit, and did no better (AUROC 0.757 versus 0.759). Adding nine cross-register ZZ correlators to the same six-qubit, eight-layer circuit raised mean AUROC to 0.767 at a cost of nine readout parameters. The eight-qubit linear classifier with 24 single-wire and 16 ZZ features reached 0.780. Because it also changed register size, this comparison did not isolate the effect of correlators at eight qubits.

Table 6 compares pair-only readouts for the same eight-qubit, eight-layer circuit under fixed training settings. ZZ features gave AUROC 0.774 with a linear classifier and 0.782 with a nonlinear

classifier; XX/YY/ZZ features gave 0.781 and 0.792, respectively. Thus richer pair measurements and a nonlinear classifier both improved the mean, with the combined design also giving the highest AUPRC (0.695) and F1 (0.642).

Table 5. Readout ablation for Qfusion, mean ± SD AUROC across the 5 family-and-clan-disjoint folds. "Single-wire obs." counted the per-wire Pauli-X/Y/Z expectation values and "Cross-register obs." the two-qubit $Z_iZ_j$ correlators between registers. All configurations used the same five evaluation partitions and the linear readout.

| Configuration | Qubits | Single-wire obs. | Cross-register obs. | Params | AUROC |
|---|---|---|---|---|---|
| 6q, 8 layers, xyz | 6 (3+3) | 18 | 0 | 185,827 | 0.759 ± 0.041 |
| 12q, 6 layers, xyz | 12 (6+6) | 36 | 0 | 195,421 | 0.757 ± 0.051 |
| 6q, 8 layers, xyz+zz | 6 (3+3) | 18 | 9 | 185,836 | 0.767 ± 0.033 |
| 8q, 8 layers, xyz+zz | 8 (4+4) | 24 | 16 | 192,233 | 0.780 ± 0.041 |

Table 6. Pair-readout comparison for the same eight-qubit, eight-layer circuit and fixed training settings (Section 2.5). All readouts omitted single-wire features. Values are mean ± sample SD over the five folds. Nonlinear denotes a 32-unit GELU classifier with dropout 0.1.

| Pair features | Classifier | AUROC | AUPRC | F1 |
|---|---|---|---|---|
| ZZ (16) | Linear | 0.774 ± 0.026 | 0.667 ± 0.103 | 0.622 ± 0.083 |
| XX/YY/ZZ (48) | Linear | 0.781 ± 0.036 | 0.674 ± 0.092 | 0.636 ± 0.068 |
| ZZ (16) | Nonlinear | 0.782 ± 0.036 | 0.675 ± 0.089 | 0.627 ± 0.070 |
| **XX/YY/ZZ (48)** | **Nonlinear** | **0.792 ± 0.038** | **0.695 ± 0.090** | **0.642 ± 0.043** |

## 4. Discussion and Conclusion

We have described a hybrid quantum-classical fusion model in which fixed weight protein and RNA embeddings were projected into an eight-qubit circuit with cross-register entangling gates. Its mean AUROC of 0.792 across disjoint RNA families and protein clans exceeded the parameter-matched control and was competitive with established methods. The ablations favored richer pair measurements and nonlinear readout over increasing circuit width or depth. Qfusion used quantum state evolution and Pauli observables, evaluated by classical statevector simulation. Standard attention as in RPIembeddor instead forms compatibility-weighted combinations of value vectors from queries and keys [8, 16]. Both represent input interactions; Qfusion constructed its joint state with unitary rotations and fixed CZ/CNOT gates.

These results suggest that an entangling coupling mechanism can serve as a practical substitute for a learned fusion operator when predicting RNA–protein interactions across families and clans absent from training. The next step is to test whether this advantage holds on larger, independently curated interaction sets and extends to screening candidate RNA sequences against a fixed protein target, the use case this method is intended to support.

## Acknowledgments

This work was supported in part by the National Science Foundation grants #IIS-1817239 and #IIS-3401096.

## References

1. M. Ramanathan, D. F. Porter, P. A. Khavari, Methods to study RNA-protein interactions, Nat. Methods 16(3), 225-234 (2019).
2. U. K. Muppirala, V. G. Honavar, D. Dobbs, Predicting RNA-protein interactions using only sequence information, BMC Bioinformatics 12, 489 (2011).
3. V. Suresh, L. Liu, D. Adjeroh, X. Zhou, RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information, Nucleic Acids Res. 43(3), 1370-1379 (2015).
4. H. Liu, Y. Jian, C. Zeng, Y. Zhao, RNA-protein interaction prediction using network-guided deep learning, Commun. Biol. 8(1), 247 (2025).
5. Y. Liu, T. Bao, P. Yin, S. Wang, Y. Wang, RPIPLM: Prediction of ncRNA-protein interaction by post-training a dual-tower pretrained biological model with supervised contrastive learning, PLoS ONE 20(8), e0329174 (2025).
6. M. Catalano, G. Pepe, G. Ausiello, C. McWhite, G. Gambosi, M. Helmer Citterich, P. F. Gherardini, Cross-Attention Over RNA and Protein Sequences Enables Generalizable Interaction Prediction, bioRxiv 2026.04.22.720174 (2026).
7. I. Shin, K. Kang, J. Kim, S. Sel, J. Choi, J.-W. Lee, H. Y. Kang, G. Song, AptaTrans: a deep neural network for predicting aptamer-protein interaction using pretrained encoders, BMC Bioinformatics 24, 447 (2023).
8. D. Matus, F. Runge, J. K. Franke, L. Gerne, M. Uhl, F. Hutter, R. Backofen, RNA-Protein Interaction Classification via Sequence Embeddings, bioRxiv 2024.11.08.622607 (2024).
9. A. Perez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum 4, 226 (2020).
10. V. Havlicek, A. D. Corcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature 567(7747), 209-212 (2019).
11. M. Schuld, N. Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. 122(4), 040504 (2019).
12. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379(6637), 1123-1130 (2023).
13. J. Chen, Z. Hu, S. Sun, Q. Tan, Y. Wang, Q. Yu, L. Zong, L. Hong, J. Xiao, T. Shen, I. King, Y. Li, Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions, arXiv:2204.00300 (2022).
14. I. Loshchilov, F. Hutter, Decoupled Weight Decay Regularization, International Conference on Learning Representations (2019), arXiv:1711.05101.
15. I. Loshchilov, F. Hutter, SGDR: Stochastic Gradient Descent with Warm Restarts, arXiv:1608.03983 (2016).
16. A. Vaswani et al., Attention Is All You Need, Advances in Neural Information Processing Systems 30 (2017), arXiv:1706.03762.