Computational and Structural Biotechnology Journal 27 (2025) 672–681

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

MMRT: MultiMut Recursive Tree for predicting functional effects of high-order protein variants from low-order variants

Bryce Forrest [a,1], Houssemeddine Derbel [a,1], Zhongming Zhao [b], Qian Liu [a,c,*]

[a] Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, 4505 S Maryland Pkwy, Las Vegas, NV 89154, USA
[b] Center for Precision Health, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[c] School of Life Sciences, College of Sciences, University of Nevada, Las Vegas, 4505 S Maryland Pkwy, Las Vegas, NV 89154, USA

ARTICLE INFO

Keywords:
Deep learning
High-order protein variants
Low-order variants
Functional effects

ABSTRACT

Protein sequences primarily determine their stability and functions. Mutations may occur at one, two, or three positions at the same time (low-order variants) or at multiple positions simultaneously (high-order variants), which affect protein functions. So far, low-order variants, such as single variants, double variants, and triple variants, have been well-studied through high-throughput experimental scanning techniques and computational prediction methods. However, research on high-order variants remains limited because of the difficulty of scanning an exponentially large number of potential variant combinations. Nonetheless, studying higher-order variants is crucial for understanding the pathogenesis of complex diseases, advancing protein engineering, and driving precision medicine. In this work, we introduce a novel deep learning model, namely *MultiMut Recursive Tree* (MMRT), to address this challenge of predicting the functional effects of high-order variants. MMRT integrates deep learning with a recursive tree framework to leverage the information from low-order variants to predict functional effects of high-order variants. We evaluated MMRT on datasets comprising 685,593 high-order variants. Our results (mean Spearman's correlation coefficient 0.55) demonstrated that MMRT outperformed three existing state-of-the-art methods: ESM (evolutionary scale modeling), DeepSequence, and ECNet (evolutionary context-integrated neural network). MMRT thus provides more accurate prediction of the functional effects of high-order protein variants, offering great potential for aiding the interpretation of variants in human disease studies.

1. Introduction

Proteins are fundamental molecular workers in cells, and their functions, stability, and interactions are primarily determined by their amino acid sequences. Although protein sequences are evolutionarily conserved, they can undergo substantial alterations that lead to protein variants. These protein variants often exhibit diverse stability and interactions, sometimes resulting in altered or novel functions. For example, spike proteins in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) acquired diverse mutations that altered infectivity, disease severity, and interaction with host immunity [1]. Similarly, seasonal influenza viruses evolve rapidly and accumulated variants continue to cause world-wide epidemics annually [2]. Protein variants are also used as cancer-specific biomarkers [3] and therapeutic targets [4] for complex human diseases. The study of protein mutations could reveal adaptive and neutral changes of cell fitness, molecular divergence, and the interplay between natural selection and genetic drift [5,6]. Knowledge of the mutational effects of protein variants can also facilitate the rational design of proteins with enhanced or novel functionalities, the optimization of enzymes for industrial applications, and the development of synthetic biology tools [7]. Therefore, a deep understanding of the functional effects of protein variants is critical to advancements in biology, evolution, medicine, and protein engineering.

Protein variants can be classified as low-order or high-order variants according to their number of mutated positions. Low-order variants, such as single, double, and triple variants, involve alterations at one, two, or three positions respectively. High-order variants are characterized by simultaneous mutations at more than three positions. Low-order variants, especially single and double variants, have been widely studied via high-throughput experimental approaches [8], such as deep

* Corresponding author at: Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, 4505 S Maryland Pkwy, Las Vegas, NV 89154, USA.
E-mail address: qian.liu@unlv.edu (Q. Liu).
[1] Equal contribution.

https://doi.org/10.1016/j.csbj.2025.02.012
Received 25 November 2024; Received in revised form 10 February 2025; Accepted 17 February 2025
Available online 18 February 2025

mutational scanning [9], yeast surface display [10], and GigaAssay [11]. These approaches have been used to comprehensively investigate dozens of proteins [11–34], with diverse measures of functional effects. Computational approaches have been proposed to extend the investigation of low-order variants to a larger number of proteins and to accurately predict functional effects [35–37] of low-order variants, especially single variants. Most of those methods, such as PolyPhen-2 [38], SIFT [39], SNAP2 [40], CADD (Combined Annotation Dependent Depletion) [41], Rep2Mut [42] and Rep2Mut-2 [43], were designed primarily to evaluate single-nucleotide variants. For example, PolyPhen-2 [38] and SIFT [39] use sequence conservation and structural features to estimate the mutational effect of single positions, whereas SNAP2 is a neural network–based method trained on single-residue substitution data. However, all these methods lack explicit modeling of combinatorial effects of multiple mutated positions. To study the functional effects of multiple mutated positions, EVmutation [44] and EVE [45] infer how combinations of mutations affect fitness by using evolutionary statistical models and residue coevolution patterns, whereas DeepSequence [46] relies on a variational autoencoder to capture higher-order dependencies. Evolutionary scale modeling (ESM) [47,48], which relies on large-scale protein language models, can also be extended to assess mutation impacts. However, these models cannot reliably capture epistasis in high-order variants.

High-order variants can substantially affect protein functions, although the mutation of each of their positions might have minimal effects. Thus, high-order variants can capture synergistic and compensatory mechanisms, and they can play a substantial role in diversifying protein stability, activity, and specificity, with the potential for the introduction of novel functions and phenotypes. Those variants are crucial to understanding the pathogenesis of complex diseases, such as cancers, where multiple mutations within oncogenes or tumor suppressor genes drive disease progression and therapeutic resistance. However, the number of high-order variants increases exponentially, compared to the number of low-order variants, making it unfeasible to use experimental approaches to comprehensively assess all possible high-order variants—even for a single protein. Thus, comprehensive studies of high-order variants are limited, consequently hindering the development of computational approaches to predict the functional effects of high-order variants.

In this work, we developed a novel approach that integrates functional effects of low-order variants to accurately predict the effects of high-order variants. Our method, which we named *MultiMut Recursive Tree* (MMRT), combines deep learning frameworks with a recursive tree to capture epistasis in high-order variants. Trained on low-order variants, MMRT was then evaluated on datasets with more than 600,000 high-order variants. Our results demonstrate that MMRT achieves higher prediction performance of functional effects of high-order variants, outperforming existing state-of-the-art methods such as ESM, DeepSequence, and ECNet (evolutionary context-integrated neural network). Thus, MMRT is a valuable complementary tool for experimental scanning approaches to reduce wet lab costs and to benefit the disease study of high-order variants of more proteins. Our tool is publicly available at https://github.com/qgenlab/MMRT.

2. Materials and methods

2.1. Datasets

To evaluate prediction performance, we downloaded and used variant datasets for 11 proteins (*i.e.*, PABP_YEAST_Fields2013 [49], HIS7_YEAST_Kondrashov2017 [50], F7YBW7_MESOW_vae [51], Tat [11], F7YBW8_MESOW_Ding_2023 [52], GCN4_YEAST_Staller_2018 [53], GFP_AEQVI_Sarkisyan_2016 [54], Q6WV12_9MAXI_Somermeyer_2022 [55], PHOT_CHLRE_Chen_2023 [56], Q8WTC7_9CNID_Somermeyer_2022 [55] and D7PM05_CLYGR_Somermeyer_2022 [55]). These proteins have varying numbers of variants

Table 1
The eleven datasets used for model training and performance assessment.

Protein Variant Dataset	Datasets (No. of Mutated Positions)	Length (AA)	Measure
PABP_YEAST_Fields2013 [49] (PABP)	1128 (1), 36522 (2)	577	Growth
HIS7_YEAST_Kondrashov2017 [50] (HIS7)	168 (1), 1475 (2), 7627 (3), 25927 (4), 60263 (5), 98869 (6), 115751 (7), 97639 (8), 57139 (9), 31270 (10 +)	552	Growth
F7YBW7_MESOW_vae [51] (F7YBW7)	37(1), 499 (2), 2798 (3), 5859 (4)	540	Fitness
HIV_Tat (Tat) [11]	1616 (1), 2161 (2)	86	Transcriptional activity
F7YBW8_MESOW_Ding_2023 (F7YBW8) [52]	86 (1), 603 (2), 2159 (3), 3240 (4), 32 (5), 60 (6), 244 (7), 653 (8), 765 (9)	94	Growth
GCN4_YEAST_Staller_2018 (GCN4) [53]	55 (1), 149 (2), 380 (3), 608 (4), 406 (5), 5 (6), 14 (7), 1 (8), 1 (9), 986 (10 +)	282	Transcriptional activation
GFP_AEQVI_Sarkisyan_2016 (GFP) [54]	12777 (1), 12336 (2), 9387 (3), 6825 (4), 4298 (5), 2526 (6), 1364 (7), 627 (8), 299 (9), 191 (10 +)	239	Fitness
Q6WV12_9MAXI_Somermeyer_2022 (CAPSD) [55]	15992 (1), 8762 (2), 3575 (3), 1300 (4), 409 (5), 159 (6), 45 (7), 9 (8), 6 (9), 3 (10 +)	223	Fitness
PHOT_CHLRE_Chen_2023 (PHOT) [56]	176 (1), 978 (2), 3565 (3), 9603 (4), 19193 (5), 29255 (6), 34230 (7), 30925 (8), 21167 (9), 16315 (10 +)	119	Fitness
Q8WTC7_9CNID_Somermeyer_2022 (Q8WTC7) [55]	11260 (1), 11641 (2), 5762 (3), 2324 (4), 857 (5), 277 (6), 98 (7), 27 (8), 20 (9), 43 (10 +)	239	Fitness
D7PM05_CLYGR_Somermeyer_2022 (D7PM05) [55]	10148 (1), 6315 (2), 3417 (3), 1752 (4), 935 (5), 427 (6), 220 (7), 76 (8), 33 (9), 23 (10 +)	236	Fitness

*XXX(Y): XXX is the number of variants, whereas Y is the number of mutated positions in a variant. Length indicates the number of amino acids (AA) in a protein sequence. Measure indicates the measurement of functional effects of variants.

(Table 1). Two proteins contained both single variants and double variants, whereas the other nine proteins contained variants with mutations at three or more positions. For example, a triple variant like "L59A:W60I:K64R" means that leucine (L) at position 59 is replaced by alanine (A), tryptophan (W) at position 60 is replaced by isoleucine (I), and lysine (K) at position 64 is replaced by arginine (R) simultaneously in the protein. Considering the number of mutated positions in a variant, we categorized and denoted the datasets as *dataset_name (n),* where *n* represents the number of mutated positions in the variant. For instance, "PABP_YEAST (2)" refers to a dataset with double variants in PABP_YEAST.

The 11 proteins consist of 859,147 variants, including 53,443 single variants, 81,441 double variants, 38,670 triple variants, 57,438 quadruple variants, and 628,155 other high-order variants. Each variant is associated with an experimental measure of functional effect upon the variants, as determined by multiplexed assays for variant effects (MAVE) [46] or GigaAssay [11]. The measurement scores served as the ground truth for model training and performance evaluation.

2.2. Methods

We proposed a novel deep learning algorithm to predict the functional effects of high-order variants when using low-order variants for training. This method holds significant promise because the number of high-order variants is exponentially larger than the number of low-order variants. For example, in a protein consisting of 100 amino acids, there are 1900 single variants, 1,786,950 double variants, and 1,109,205,300 triple variants, but the number of quadruple variants escalates to 511,017,963,225. The triple variants represent 0.21 % of quadruple variants. Thus, although it is feasible to use experimental scanning approaches to experimentally determine the functional effects of single and double variants, as well as of a portion of triple variants [11,46], it is impractical to perform such scanning for quadruple variants or even for high-order variants because of the sheer volume of possible combinations. We therefore sought to develop a deep learning algorithm, MMRT, to address this challenge.

MMRT uses a recursive tree structure [57] to integrate mutation effects from multiple mutated positions in a variant. The input to MMRT is a protein sequence and its variant, and the output is the predicted functional effect of a high-order variant (denoted as $v_i = \{p_0, p_1, \ldots p_k, \}$, where k positions are mutated simultaneously).

2.2.1. Vector generation

To process the variants, we used a protein sequence as input for ESM [47,48] to generate a representation vector of each position in the input protein. ESM is a self-supervised learning approach composed of multiple transformer layers and has been trained on millions of protein sequences using the masked language modeling objective. ESM endeavors to learn multiple levels of protein knowledge such as biochemical properties and evolutionary information. ESM has several different releases, and we used esm1v_t33_650M_UR90S_1 and refer it to as ESM_r to distinguish it from the other ESM methods for functional effect prediction. With ESM_r, each mutated position is represented by a 1280-dimensional vector that captures the contextual information hidden in the protein sequence. The representation of individual residues has been demonstrated to benefit various protein tasks (e.g., secondary structure prediction [58] and solvent accessibility prediction [59]), and we extended the benefits to predict mutational effects of high-order variants.

During training processes, ESM_r was frozen and recursive neural network layers were trained. The representations generated via ESM_r were used as inputs of MMRT in both training and testing processes, although training processes and testing processes have different iterations of representation integration. That is, training processes on low-order variants usually have one or two iterations of representation integration, whereas testing processes on higher-order variants require several iterations. For example, MMRT with a triple variant as input has two iterations of representation integration (Fig. 1), whereas MMRT with a 10-position variant as input has nine recursive integrations.

2.2.2. Recursive integration

After vector generation, the 1280-dimensional vector of each mutated position was converted to a 64-dimensional vector by a fully connected layer. For a variant $v_i = \{p_0, p_1, \ldots p_k\}$, $Q_i = \{q_0, q_1, \ldots q_k\}$ is used to represent the converted vectors, where q_i is a 64-dimensional vector for position p_i. We applied a recursive tree structure [57] to integrate information from each mutated position with $k-1$ iterations. In each iteration $j > 0$, we concatenated q_{j-1} and q_j via a fully connected layer to produce a 64-dimensional vector that replaced the old vector q_j. We repeated this concatenation until all vectors in Q_i were merged into a single 64-dimensional vector (i.e., $j = k$), as demonstrated in Fig. 1 for a variant with three mutated positions.

2.2.3. Vector integration for mutated sequence

We generated a mutated protein sequence from the input protein sequence by replacing all position p_i with the corresponding mutations from the variant v_i. We used the mutated sequence as input for ESM_r and generated representation vectors for each mutated position. As with Q_i, we converted them into a 64-dimensional vector, that is, $Q_i' = \{q_0', q_1', \ldots q_k'\}$. We then merged all q_j' to a single 64-dimensional vector using the same recursive concatenation process described above (Fig. 1).

2.2.4. Final prediction

The two final 64-dimensional vectors, one from Q_i and one from Q_i', were concatenated through a fully connected layer to generate a 64-dimensional vector. This final vector was used as input for a fully connected layer with a sigmoid activation function to predict the functional effect of the variant v_i.

We also designed a window-based version of MMRT, called *MMRT-WIN,* to incorporate the context around the mutated positions. In MMRT-WIN, instead of representation vectors of single positions, the q_j vector encompasses several consecutive positions centered around the mutated positions. In this study, we used one upstream and one downstream position relative to each mutated position. The recursive concatenation process of MMRT-WIN follows the same iterative procedure that was used in MMRT to integrate a broader sequence context into the prediction model.

2.3. Training and fine-tuning

We used two different strategies to train and test MMRT and MMRT-WIN: the blind-variant testing and few-shot learning. In the blind-variant testing strategy, we trained MMRT(-WIN) on low-order variants, and then tested its performance on high-order variants. For example, MMRT(-WIN) was trained on low-order variants (single, double, and triple variants) of a protein, and then was used to predict functional effects of high-order variants. To train a model on low-order variants, we set the number of epochs to 2000, the batch size to 32, and the learning rate to 1e-06. We used the Adam optimizer [60] and mean squared error (MSE) loss function to guide the back-propagation process [61]. MSE is defined in Eq. (1) where n is the number of variants, Y_i is the number of observed measurements, and $\widehat{Y}$ is the number of predicted measurements.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 \quad (1)$$

Trained blind-variant testing models were then fine-tuned to create few-shot models. During fine-tuning, we trained a blind-variant testing model on 10 % of high-order variants with a batch size of 32 and a learning rate of 5e-07 over 4000 epochs. We used the remaining 90 % of high-order variants to test the performance of the few-shot models. To

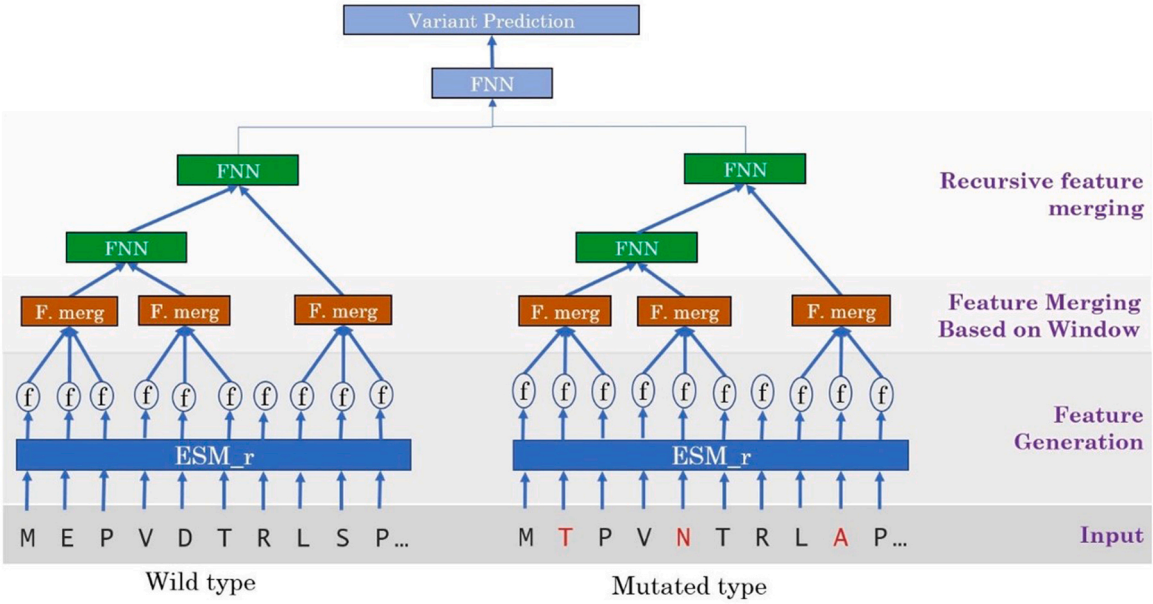

Fig. 1. Demo architecture of MultiMut Recursive Tree to predict functional effects for variants with three mutated positions. Abbreviations: ESM_r, evolutionary scale modeling (ESM) used as feature extractor; F. merg, a vector merging layer; FNN, fully connected neural network.

compare the performance of MMRT few-shot against that of three existing state-of-the-art methods, we used a 10-fold cross-validation approach. We randomly divided the dataset into 10 subsets, each containing 10 % of the experimental data. For each iteration, one subset was designated for training and the remaining nine subsets were designated for testing. This process was repeated five times, and the mean performance across these iterations was calculated to assess MMRT few-shot models.

2.4. Performance comparison

We compared our methods with several state-of-the-art methods, including ESM [47,48], DeepSequence [62], and ECNet [63]. We focus here on the comparative results with these three state-of-the-art models.

ESM [47,48] is a pretrained BERT [64] model that utilizes a transformer model to uncover complicated information hidden in protein sequences. ESM deploys a self-supervised learning technique to predict masked amino acids in a protein sequence by their surrounding context. Its learned representation of proteins can be used in various protein tasks, including the estimation of mutated effects of protein variants by calculating the logarithmic ratio of the probability between the mutated amino acid and the wild-type amino acid. We compared the performance of MMRT(-WIN) against two different ESM models: ESM-v1_t34_670M_UR50S (ESM1) and ESM-v2_t36_3B_UR50D (ESM2).

DeepSequence is a deep latent-variable model that leverages variational autoencoders [62] to extract latent factors from protein or RNA sequences, capturing higher-order correlations within biological sequence families. Trained on multiple sequence alignment (MSA) of homologous protein sequences, DeepSequence predicts the functional or stability effects of mutations through fitness scores by calculating log-likelihood differences of mutated proteins and wild-type proteins. However, DeepSequence is primarily a generative model, and it requires additional training to be able to predict the effect of a variant on each protein sequence. To train DeepSequence, we used an MSA tool to create MSA sequences for each input protein sequence. Upon a suggestion from DeepSequence, we utilized EVcoupling [65] from the website v2.evcouplings.org, with a bit score threshold of 0.5 bits/residue to generate sequences for PABP, HIS7, F7YBW7, and Tat. For the proteins F7YBW8, GCN4, GFP, CAPSD, PHOT, Q8WTC7, and D7PM05, we used MSA files provided by ProteinGym [66]. These MSA sequences were subsequently used to retrain DeepSequence for precise mutational effect predictions.

The deep learning algorithm ECNet [63] combines local and global evolutionary representations to predict functional fitness. The global evolutionary representation is generated by a language model trained with self-supervised learning on comprehensive semantic-rich global sequences from protein sequence databases such as UniProt [67] and Pfam [68]. The local evolutionary coupling is derived by fitting a direct coupling analysis model to homologous MSA sequences of a protein to generate a coupling matrix. The matrix quantifies the co-constraints of all possible 20^2 amino acid combinations between two positions in the sequence. The global and local representations are then fed into a bidirectional long short-term memory network (BiLSTM) to predict protein fitness. ECNet provides two implementations: ECNet-base and ECNet-ensemble. ECNet-base models contain three independent predictors, each with a different loss [63]. Since ECNet does not provide well-trained models for testing, we used the default parameters recommended by ECNet to train three ECNet-base models with 100 epochs for performance evaluation: one ECNet-base model was trained on single variants, one was trained on single and double variants, and one was trained on single, double, and triple variants. In addition, we also trained an ECNet-ensemble model, which consists of five replicates of ECNet-base models, each with a different loss objective [63]. The predicted fitness of a variant from the ECNet-ensemble was averaged across the five replicates. These ECNet-base and ECNet-ensemble models were then used to predict the functional effects of higher-order variants.

2.5. Evaluation measurements

We used Spearman's rank correlation coefficient (SRCC) to evaluate the prediction performance generated by each method. SRCCs were calculated using the Python package SciPy between the predicted and experimental estimates, as shown in Eq. (2) below.

$$SRCC = \frac{\mathrm{cov}(R(X), \quad R(Y))}{\sigma_{R(X)} \quad \sigma_{R(Y)}} \tag{2}$$

Where X and Y are the experimental and predicted estimates, respectively, for a set of variants. SRCC, ranging from −1 to 1, provides a measure of the association between experimental and predicted

estimates.

3. Results

3.1. Predicting functional effects of higher-order variants using blind-variant testing models

We first evaluated the performance of our MMRT-trained models in predicting the functional effects of high-order variants. Three different MMRT models were trained on low-order variants under a blind-variant testing strategy: (1) an MMRT-TR123 model was trained on single, double, and triple variants, (2) an MMRT-TR123-WIN model was trained on single, double, and triple variants, with a window size of three (one amino acid before and one after), and (3) an MMRT-TR23-WIN model was trained on double and triple variants, with a window size of three. We assessed the prediction performance of MMRT on more than 600,000 high-order variants of nine proteins, and compared its performance with the performance of several state-of-the-art methods (i.e., ESM, DeepSequence, and two ECNet models) (Tables 2 and 3; Figs. 2 and 3).

As shown in Table 2 and Fig. 2, the MMRT-TR23-WIN model outperformed ESM1 on 25 datasets of 30 datasets with different numbers of mutated positions, with an improvement in SRCC ranging from 0.04 on F7YBW7(4) to 0.87 on Q6WV12 (4), where XXXX(Y): XXX is the protein name and Y is the number of mutated positions in higher-order variants. ESM-1 outperformed MMRT-TR23-WIN on only 5 datasets, with an improvement in SRCC ranging from 0.05 on GCN4 (5) to 0.17 on GCN4 (4). On average, MMRT-TR23-WIN achieved an overall improvement of 0.29 over ESM1.

MMRT-TR23-WIN also surpassed ESM2 in 21 of 30 datasets, with an improvement in mean SRCC of 0.27. The SRCC improvement ranged from 0.04 on GFP(8 +) to 0.88 on Q6WV12 (4). For instance, on HIS7 (4), ESM1 and ESM2 achieved an SRCC of 0.31 and 0.34, respectively, while MMRT-TR23-WIN achieved a significantly higher correlation of SRCC 0.62. Similarly, on GFP (4), ESM1 and ESM2 yielded an SRCC of 0.14 and 0.01, respectively, whereas MMRT-TR23-WIN achieved a superior performance of SRCC 0.78. Thus, MMRT-WIN outperformed the ESM models.

The MMRT-TR23-WIN model also outperformed DeepSequence across all 30 datasets. MMRT-TR23-WIN achieved a mean SRCC of 0.55, surpassing the SRCC of 0.19 achieved by DeepSequence. The improvement varied from SRCC 0.03 on F7YBW7 (4) to SRCC 0.80 on Q8WTC7 (4).

Compared with ECNet, the MMRT-TR23-WIN model outperformed ECNet-base-TR123 on 19 of 23 datasets, while giving similar results on 1 dataset (PHOT was not used, because its high-order variants were generated using ECNet and MAVE-NN). The improvement in SRCC ranged from 0.01 on HIS7 (4,5) and D7PM05 (4) to 0.16 on D7PM05 (5). Compared with ECNet-ensemble-TR123, the MMRT-TR23-WIN model provided better results in 16 of 23 datasets, with SRCC improvement ranging from 0.01 on D7PM05(4, 5 +), and GFP (6, 8 +) to 0.08 on HIS7 (10 +), while ECNet-ensemble-TR123 produced better results in 6 datasets, with SRCC improvement ranging from 0.01 on HIS7(4), F7YBW8(4), GFP (5), and Q8WTC7 (4) to 0.13 on GCN (4).

In summary, our evaluations demonstrated that MMRT-WIN, trained on lower-order variants, achieves excellent performance in predicting the functional effects of higher-order variants. MMRT outperformed the existing state-of-the-art methods and achieved a mean SRCC of 0.55. In comparison, the mean SRCCs for ESM1, ESM2, DeepSequence, ECNet-ensemble-TR123, and ECNet-base-TR123 were 0.26, 0.28, 0.19, 0.54, and 0.52, respectively.

Table 2
Performance prediction for quadruple or higher-order variants with MMRT(-WIN) compared to state-of-the-art models. Performance is evaluated using Spearman's rank correlation coefficient (SRCC).

Protein	#Position in variant	ESM1	ESM2	DeepSequence	ECNet-ensemble-TR123	ECNet-base-TR123	MMRT-TR123	MMRT-TR123-WIN	MMRT-TR23-WIN
F7YBW7	4	0.11	**0.16**	0.12	0.13	0.11	**0.16**	0.11	0.15
HIS7_2017	4	0.31	0.34	0.36	**0.63**	0.61	0.61	0.62	0.62
	5	0.32	0.37	0.39	**0.67**	0.66	0.64	**0.67**	**0.67**
	6	0.31	0.39	0.40	0.68	0.67	0.66	**0.70**	**0.70**
	7	0.3	0.41	0.40	0.67	0.66	0.66	**0.71**	**0.71**
	8	0.32	0.43	0.43	0.66	0.65	0.65	0.70	**0.71**
	9	0.38	0.49	0.50	0.64	0.63	0.64	**0.70**	**0.70**
	10 +	0.55	0.62	0.63	0.59	0.56	0.63	**0.67**	**0.67**
Q6WV12	4	0.01	0	0.19	0.85	0.85	0.84	**0.88**	**0.88**
	5 +	0.08	0.07	0.30	0.77	0.77	0.78	0.81	**0.82**
Q8WTC7	4	0.03	0.08	0.05	0.86	0.85	**0.88**	0.85	0.85
	5 +	0.02	0.01	0.08	0.79	0.79	0.75	0.82	**0.83**
GCN4	4	0.3	**0.29**	0.02	0.26	0.21	0.14	0.27	0.13
	5	**0.15**	**0.15**	0.08	0.06	0.03	0.13	0.06	0.10
D7PM05	4	0.08	0.03	0.12	0.68	0.68	**0.70**	0.69	0.69
	5 +	0.02	0.04	0.09	0.26	0.11	**0.31**	0.28	0.27
GFP	4	0.14	0.01	0.08	0.81	0.81	0.79	**0.82**	0.78
	5	0.08	0.01	0.08	0.66	0.64	0.65	**0.67**	0.65
	6	0.05	0.01	0.04	0.40	0.40	0.42	**0.45**	0.41
	7	0	0	0	0.2	0.21	**0.23**	**0.23**	**0.23**
	8 +	0.06	0.05	0	0.08	0.01	0.09	0.08	**0.09**
F7YBW8	4	0.32	0	0.12	0.67	0.70	0.70	**0.71**	0.66
	5 +	0.35	0.09	0	0.46	0.40	0.47	0.44	**0.48**
PHOT	4	0.59	0.65	NA	NA	NA	0.63	0.62	**0.65**
	5	0.57	0.66	NA	NA	NA	0.64	**0.76**	**0.76**
	6	0.55	**0.65**	NA	NA	NA	0.32	0.39	0.44
	7	0.52	**0.64**	NA	NA	NA	0.42	0.62	0.62
	8	0.50	**0.64**	NA	NA	NA	0.17	0.28	0.33
	9	0.47	**0.63**	NA	NA	NA	0.32	0.53	0.56
	10 +	0.43	**0.55**	NA	NA	NA	0.15	0.27	0.36

Abbreviations: ECNet, evolutionary scale modeling; ESM, evolutionary scale modeling; MMRT, MultiMut Recursive Tree; NA, not available. Bolded SRCCs indicate the highest performance for a row. The TR123 model was trained on single, double, and triple variants, the TR23 model was trained on double and triple variants, and the WIN model was trained using a window of one position upstream and one position downstream of mutated positions. On PHOT, a DeepSequence run cannot be completed after 15 days, and ECNet was used to generate the variant effects; thus, the blind-variant testing results are not available (NA) for both methods on PHOT.

Table 3
Performance of the MultiMut Recursive Tree (MMRT) few-shot model fine-tuned on 10 % of the data. Performance is evaluated using Spearman's rank correlation coefficient (SRCC).

Protein	#Position in Variant	ESM1	ESM2	DeepSequence	ECNet-ensemble-TR123	MMRT-TR123-WIN	MMRT Few-shot
tat	2	0.16	0.04	0.10	-	-	**0.38**
PABP	2	0.7	0.68	0.71	-	-	**0.87**
F7YBW7	2	0.39	0.11	0.67	-	-	**0.68**
	3	0.47	0.27	0.52	-	-	**0.64**
	4	0.11	**0.16**	0.12	0.12	0.11	0.14
HIS7_2017	4	0.31	0.34	0.36	**0.63**	0.62	0.62
	5	0.32	0.37	0.39	0.67	0.67	**0.71**
	6	0.31	0.39	0.40	0.68	0.70	**0.79**
	7	0.30	0.41	0.40	0.67	0.71	**0.84**
	8	0.32	0.43	0.43	0.66	0.70	**0.86**
	9	0.38	0.49	0.50	0.64	0.70	**0.87**
	10 +	0.55	0.62	0.63	0.59	0.67	**0.82**
Q6WV12	4	0.01	0	0.19	0.85	0.84	**0.89**
	5 +	0.08	0.07	0.30	0.77	0.75	**0.81**
Q8WTC7	4	0.03	0.08	0.05	**0.86**	0.85	0.85
	5 +	0.02	0.01	0.08	0.79	0.74	**0.86**
GCN4	4	0.30	0.29	0.02	**0.26**	0.27	0.15
	5	0.15	0.15	0.08	0.06	0.06	**0.18**
D7PM05	4	0.08	0.03	0.12	0.68	0.41	**0.67**
	5 +	0.02	0.04	0.09	0.26	0.15	**0.28**
GFP	4	0.14	0.01	0.08	0.81	0.67	**0.81**
	5	0.08	0.01	0.08	0.66	0.45	**0.69**
	6	0.05	0.01	0.04	0.40	0.23	**0.44**
	7	0	0	0	0.2	0.1	**0.37**
	8 +	0.06	0.05	0	**0.08**	0.03	0.04
F7YBW8	4	0.32	0	0.12	0.67	0.65	**0.68**
	5 +	0.35	0.09	0	0.08	0.07	**0.48**
PHOT	4	0.59	0.65	NA	NA	0.76	**0.93**
	5	0.57	0.66	NA	NA	0.39	**0.91**
	6	0.55	0.65	NA	NA	0.62	**0.94**
	7	0.52	0.64	NA	NA	0.28	**0.89**
	8	0.50	0.64	NA	NA	0.53	**0.94**
	9	0.47	0.63	NA	NA	0.23	**0.89**
	10 +	0.43	0.55	NA	NA	0.30	**0.91**

Abbreviations: ECNet, evolutionary context-integrated neural network; ESM, evolutionary scale modeling; NA, not available. For ECNet, the best results from Table 2 were selected. The ESM and DeepSequence results are the same as their results listed in Table 2 and included here for a quick reference. Bolded SRCCs indicate the highest performance for a row.

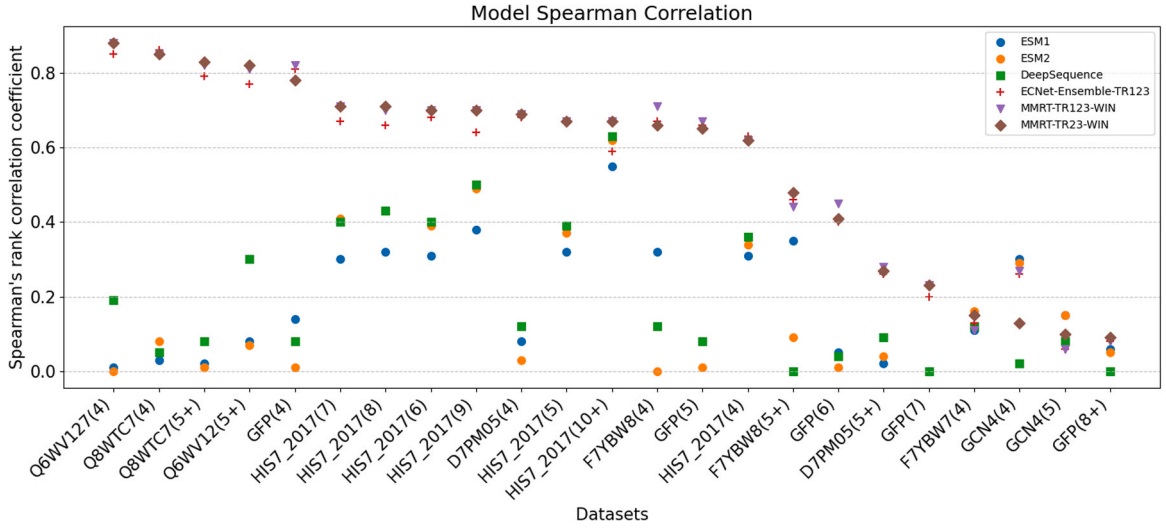

Fig. 2. The performance of MMRT-TR123-WIN and MMRT-TR23-WIN compared with existing state-of-the-art methods, as measured by Spearman's rank correlation coefficient. XXXX(Y): XXX is the protein name, and Y is the number of mutated positions in higher-order variants. Abbreviations: ECNet, evolutionary context-integrated neural network; ESM, evolutionary scale modeling; MMRT, MultiMut Recursive Tree.

3.2. Performance of various MMRT models

We subsequently tested the performance of MMRT models trained on single, double, and triple variants with and without window setting, that is, the MMRT-TR123-WIN and MMRT-TR123 models, respectively. As shown in Table 2, MMRT-TR23-WIN outperformed MMRT-TR123 on 18 of 30 datasets, and it outperformed MMRT-TR123-WIN on 12 datasets, while producing similar results on 12 datasets. This comparison indicates that single variants do not contribute to the capture of epistatic interactions that are crucial for accurate predictions of the functional

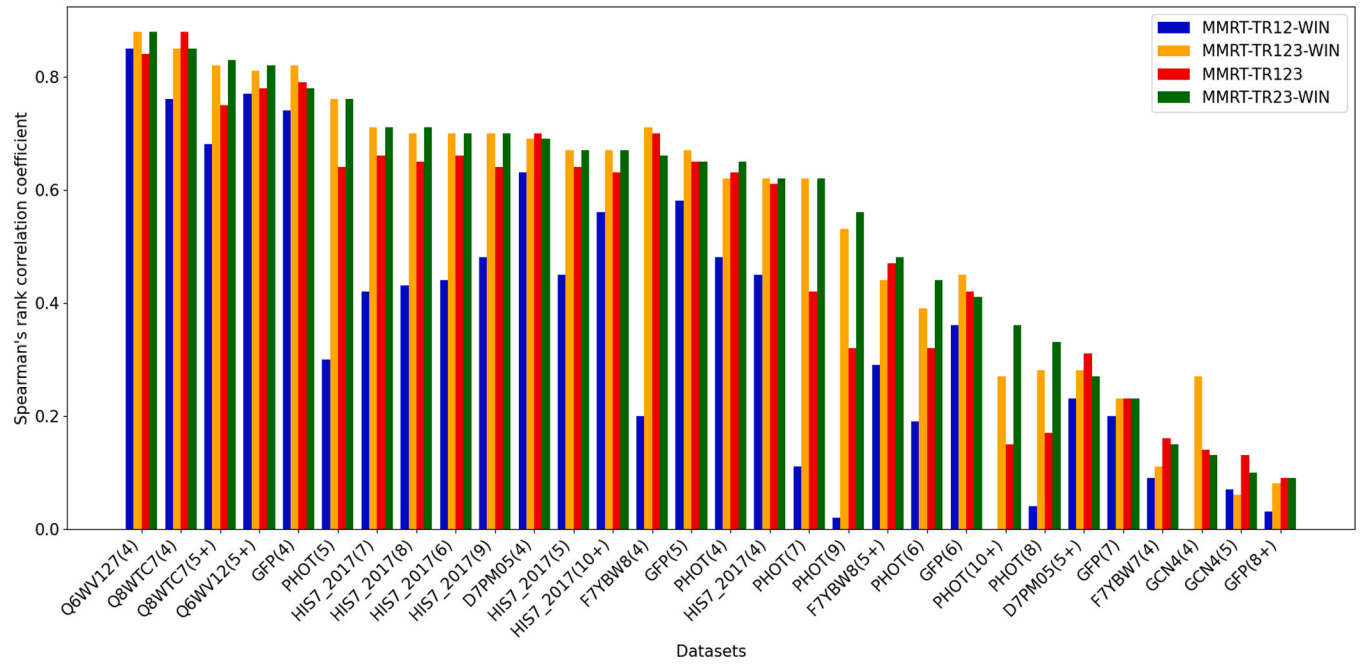

Fig. 3. Comparative performances of the four different MultiMut Recursive Tree (MMRT) models measured by Spearman's rank correlation coefficient. TR12-WIN was trained on single and double variants, TR123-WIN was trained on single, double, and triple variants, and TR23-WIN was trained on double and triple variants, and all three of these models were trained using a window of one position before and one position after each mutated position. TR123 was trained with only the mutated position used as input. XXXX(Y): XXX is the protein name, and Y is the number of mutated positions in higher-order variants.

effects of higher-order variants. Thus, it is better to include both double and triple variants to capture epistasis of mutated positions.

Nevertheless, MMRT-TR123-WIN also outperformed ESM1 in 24 of the 30 datasets, with SRCC improvements ranging from 0.02 on GFP (8 +) to 0.87 on Q6WV12 (4). The improvement in mean SRCC over ESM1 was 0.29. MMRT-TR123-WIN surpassed ESM2 in 21 out of 30 datasets, with an improvement in mean SRCC of 0.26, and it outperformed DeepSequence on 21 of 23 datasets. When compared to ECNet, MMRT-TR123-WIN generated better performance than ECNet-TR123, and it generated competitive results compared to ECNet-Ensemble-TR123. In summary, the MMRT-TR123-WIN model outperformed all three existing state-of-the-art models: ESM, DeepSequence, and ECNet. The MMRT-TR123-WIN model had a mean SRCC of 0.54, compared to 0.26 for ESM1, 0.28 for ESM2, 0.19 for DeepSequence, and 0.47 for ECNet-TR123.

We also compared the performance of MMRT trained only on single and double variants (MMRT-TR12-WIN1) with the performance of MMRT-TR-123-WIN and MMRT-TR-23-WIN. As summarized in Fig. 3, the results showed that the performance of the model decreases without the inclusion of triple variants, which indicates the necessity of including triple variants in the training of MMRT models.

3.3. Predicting functional effects of higher-order variants using few-shot models

We further tested the performance of MMRT few-shot models, which were fine-tuned on 10 % of the higher-order variants and tested on the remaining 90 %. To avoid bias from random splits, we repeated this fine-tuning process 10 times, and averaged the performance across the 10 trials, as presented in Table 3 and Fig. 4. The results clearly demonstrated that our fine-tuned MMRT models significantly improved the prediction of functional effects of high-order variants for most datasets. MMRT consistently outperformed all three existing state-of-the-art methods on 25 of 30 tested datasets. This significant improvement highlights the capability of MMRT to effectively leverage a small subset of high-order variants to make accurate predictions for a vast number of

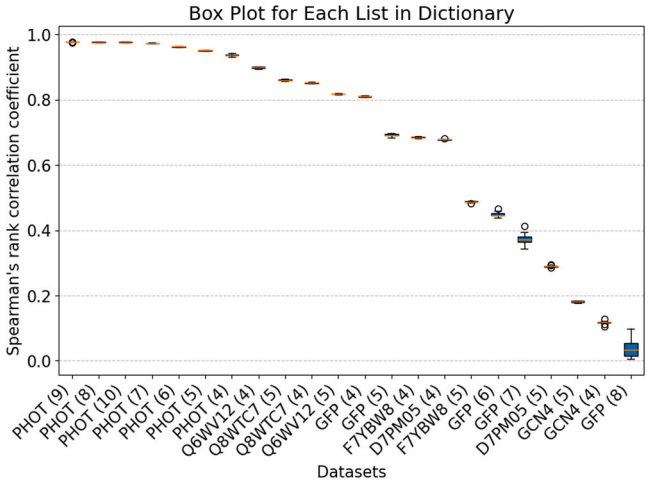

Fig. 4. The performance distribution of the MultiMut Recursive Tree few-shot models, as measured by Spearman's rank correlation coefficient. Each box corresponds to a dataset, representing the variability in performance across the 10 models trained on that dataset. Error bars indicate the standard deviation of the performance of the 10 models. XXXX(Y): XXX is the protein name, and Y is the number of mutated positions in higher-order variants.

unseen high-order variants. Our fine-tuning of MMRT on limited high-order variants showcases the potential for combining MMRT with minimal experimental variant scanning to accurately estimate the functional effects of millions of high-order variants, which would offer an innovative and practical solution to the challenges posed by the sheer volume of high-order variants.

4. Discussion

In this research, we developed a novel approach, MMRT, to predict the functional effects of high-order variants by leveraging lower-order

variants. MMRT was evaluated on 685,593 high-order variants, and the assessment across different evaluation strategies demonstrated that MMRT outperformed existing methods in predicting the functional effects of high-order variants. This capability is critically important, because it is particularly challenging to experimentally scan all possible high-order variants for proteins with hundreds of amino acids. MMRT effectively reduces the cost of estimating functional effects of high-order variants by leveraging experimental results from low-order variants, thus offering an efficient and practical solution for studying variations in viral and bacterial proteins.

Notably, MMRT used representation vectors generated from an ESM framework as input, but it significantly outperformed ESM alone. This improvement occurred because single variants have a limited capacity to capture the epistatic interactions inherent in high-order variants, as demonstrated by results from MMRT and ECNet models trained solely on single variants: neither model produced substantial improvement of functional effect prediction of high-order variants (Table 4). In contrast, the models trained on double and triple variants achieved much higher prediction performance. Including single variants in the training process did not substantially alter the performance of MMRT(-WIN), with a small deviation of ± 0.03. Thus, single variants alone may not be useful for predicting the functional effects of high-order variants, and complicated epistasis is common in high-order variants. We further trained MMRT on low-order variants with as many as three mutated positions to guarantee enough training variants for the learning process and to capture epistasis between mutated positions. Since the experimental cost to generate protein variants with more than three mutated positions increases exponentially compared to the cost to generate double and triple variants, our training strategy is a trade-off between learning capability and experimental cost.

We also analyzed the prediction variability of MMRT23-WIN with HIS7 (Fig. 5). For this dataset, we classified a prediction as "good" if the absolute difference between the experimental and predicted values was less than 0.4, as previously recommended [50]). The clusters of data points on the scale indicate that predictive accuracy is low for positions involved in an insufficient number of variants. This finding suggests that MMRT performs more effectively on a diverse variant set where each position is involved in a sufficient number (such as >100) of variants.

However, our method has some limitations. First, we sequentially combined multiple positions in high-order variants. Other combination strategies, such as domain-based combination, might generate better predictions, but have not been explored because there are numerous alternative combination structures for higher-order variants and various high-order variants might have different optimal combination structures. The number of higher-order variants currently available is not large enough to test possible combinations. Second, blind-variant testing models are currently protein-specific, because there are limited number of proteins which contain high-order variants with the same measurement of functional effects. As more protein datasets become available, we will be able to develop generalized blind-variant testing models applicable across multiple proteins. On those datasets, transfer learning can be used to improve the generalizability of MMRT across proteins. Third, this evaluation was based on measurements of functional effects generated by MAVE, and the reliability of MAVE might affect prediction performance. For example, all methods performed comparatively worse on the F7YBW7_MESOW_vae dataset, which warrants further investigation as more protein variant datasets become available. Nevertheless, these limitations will be overcome with the availability of more datasets.

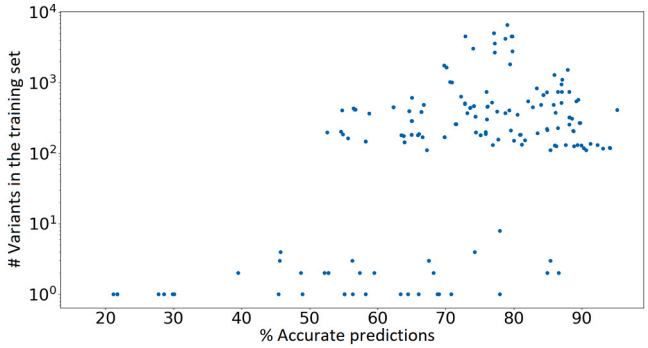

Fig. 5. Logarithmic scale indicates the accuracy of MMRT23-WIN predictions by the number of variants for HIS7. X-axis: the percentage of accurate predictions for each position; Y-axis: the number of variants in the training set. MMRT, MultiMut Recursive Tree; WIN, trained using a window of one position upstream and one position downstream of mutated positions.

5. Conclusion

In this study, we developed and evaluated MMRT, a novel architecture based on a recursive neural network and deep learning that was designed to predict functional effects of high-order variants. We compared the performance of MMRT with that of existing models (i.e., ESM, DeepSequence, and ECNet), and found that MMRT performed much better when it was trained primarily on double and triple variants. These results suggest that MMRT can substantially enhance the investigation of high-order variants and it has the potential to be of benefit to variant studies for complex human diseases.

Statement

During the preparation of this work the author(s) used ChatGPT in order to correct grammar errors. After using this tool/service, the author (s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Table 4
Performance (Spearman's rank correlation) of MMRT models trained on single variants compared with ESM1, ESM2, DeepSequence, and ECNet.

Dataset	Tat (2)	PABP (2)	F7YBW7 (2)	F7YBW7 (3)	F7YBW7 (4)	HIS7 (2)	HIS7 (3)	HIS7 (4)	HIS7 (5)	HIS7 (6)	HIS7 (7)	HIS7 (8)	HIS7 (9)	HIS7 (10 +)
ESM1	0.16	0.7	0.39	0.47	0.11	0.15	0.28	0.31	0.32	0.31	0.3	0.32	0.38	0.55
ESM2	0.04	0.68	0.11	0.27	**0.16**	0.13	0.27	0.34	0.37	0.39	0.41	0.43	0.49	0.62
DeepSequence	0.1	0.71	0.67	**0.52**	0.12	0.17	**0.3**	**0.36**	**0.39**	**0.4**	**0.4**	**0.43**	**0.5**	**0.63**
ECNet-ensemble TR1	0.25	**0.77**	0.66	0.51	0.06	0.22	0.25	0.29	0.27	0.24	0.21	0.18	0.15	0.1
ECNet-base TR1	0.26	0.75	0.64	0.48	0.05	0.17	0.22	0.23	0.2	0.14	0.1	0.09	0.09	0.11
MMRT TR1	0.34	0.65	0.54	0.4	0.14	**0.25**	0.25	0.28	0.28	0.29	0.3	0.34	0.4	0.51
MMRT TR1 (WIN)	**0.41**	0.62	**0.76**	0.47	0.12	**0.25**	0.29	0.34	0.36	0.37	0.37	0.41	0.45	0.51

Abbreviations: ECNet, evolutionary context-integrated neural network; ESM, evolutionary scale modeling; MMRT, MultiMut Recursive Tree. With TR1, the MMRT model was trained only on single mutation datasets. With WIN, MMRT was trained using a window of one position upstream and one position downstream of mutated positions.

CRediT authorship contribution statement

Liu Qian: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zhao Zhongming:** Writing – review & editing, Writing – original draft, Validation. **Derbel Houssemeddine:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Forrest Bryce:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation.

Declaration of Competing Interest

The authors claim that there is no conflict of interest.

Acknowledgements

The authors would like to thank their laboratory members for useful discussion and the reviewers for their valuable comments. This research was in part funded by the National Institute of General Medical Sciences (P20GM121325). Z.Z. was partially supported by funding from the National Institutes of Health (R01LM012806, R01LM012806-07S1, and U01AG079847) and the Cancer Prevention and Research Institute of Texas (CPRIT RP180734 and RP240610). The funders had no role in the study design, data collection and analysis, the decision to publish, or the preparation of this manuscript.

References

[1] Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 2021;19:409–24. https://doi.org/10.1038/s41579-021-00573-0.
[2] Petrova VN, Russell CA. The evolution of seasonal influenza viruses. Nat Rev Microbiol 2018;16:47–60. https://doi.org/10.1038/nrmicro.2017.118.
[3] Wang Q, Chaerkady R, Wu J, Hwang HJ, Papadopoulos N, Kopelovich L, et al. Mutant proteins as cancer-specific biomarkers. Proc Natl Acad Sci USA 2011;108:2444–9. https://doi.org/10.1073/pnas.1019203108.
[4] Xie X, Yu T, Li X, Zhang N, Foster LJ, Peng C, et al. Recent advances in targeting the "Undruggable" proteins: from drug discovery to clinical trials. Signal Transduct Target Ther 2023;8:335. https://doi.org/10.1038/s41392-023-01589-z.
[5] Jayaraman V, Toledo-Patiño S, Noda-García L, Laurino P. Mechanisms of protein evolution. Protein Sci 2022;31:e4362. https://doi.org/10.1002/pro.4362.
[6] Hershberg R. Mutation-the engine of evolution: studying mutation and its role in the evolution of bacteria: Fig. 1. Cold Spring Harb Perspect Biol 2015;7:a018077. https://doi.org/10.1101/cshperspect.a018077.
[7] Li C, Zhang R, Wang J, Wilson LM, Yan Y. Protein engineering for improving and diversifying natural product biosynthesis. Trends Biotechnol 2020;38:729–44. https://doi.org/10.1016/j.tibtech.2019.12.008.
[8] Giver L, Gershenson A, Freskgard P-O, Arnold FH. Directed evolution of a thermostable esterase. Proc Natl Acad Sci 1998;95:12809–13. https://doi.org/10.1073/pnas.95.22.12809.
[9] Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods 2014;11:801–7. https://doi.org/10.1038/nmeth.3027.
[10] Gai SA, Wittrup KD. Yeast surface display for protein engineering and characterization. Curr Opin Struct Biol 2007;17:467–73. https://doi.org/10.1016/j.sbi.2007.08.012.
[11] Benjamin R, Giacoletto CJ, FitzHugh ZT, Eames D, Buczek L, Wu X, et al. GigaAssay – An adaptable high-throughput saturation mutagenesis assay platform. Genomics 2022;114:110439. https://doi.org/10.1016/j.ygeno.2022.110439.
[12] Doud MB, Bloom JD. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. Viruses 2016;8. https://doi.org/10.3390/v8060155.
[13] Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, et al. Functional constraint profiling of a viral protein reveals discordance of evolutionary conservation and functionality. PLOS Genet 2015;11:e1005310. https://doi.org/10.1371/journal.pgen.1005310.
[14] Qi H, Olson CA, Wu NC, Ke R, Loverdo C, Chu V, et al. A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis c viral fitness and drug sensitivity. PLOS Pathog 2014;10:e1004064. https://doi.org/10.1371/journal.ppat.1004064.
[15] Mishra P, Flynn JM, Starr TN, Bolon DNA. Systematic mutant analyses elucidate general and client-specific aspects of Hsp90 function. Cell Rep 2016;15:588–98. https://doi.org/10.1016/j.celrep.2016.03.046.
[16] Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, et al. Capturing the mutational landscape of the beta-lactamase TEM-1. Proc Natl Acad Sci 2013;110:13067–72. https://doi.org/10.1073/pnas.1215206110.
[17] Firnberg E, Labonte JW, Gray JJ, Ostermeier M. A comprehensive, high-resolution map of a gene's fitness landscape. Mol Biol Evol 2014;31:1581–92. https://doi.org/10.1093/molbev/msu081.
[18] Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell 2015;160:882–92. https://doi.org/10.1016/j.cell.2015.01.035.
[19] Deng Z, Huang W, Bakkalbasi E, Brown NG, Adamski CJ, Rice K, et al. Deep sequencing of systematic combinatorial libraries reveals β-lactamase sequence constraints at high resolution. J Mol Biol 2012;424:150–67. https://doi.org/10.1016/j.jmb.2012.09.014.
[20] Rockah-Shmuel L, Tóth-Petróczy Á, Tawfik DS. Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. PLOS Comput Biol 2015;11:e1004421. https://doi.org/10.1371/journal.pcbi.1004421.
[21] Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, et al. Massively parallel functional analysis of BRCA1 RING domain variants. Genetics 2015;200:413–22. https://doi.org/10.1534/genetics.115.175802.
[22] Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci 2012;109:16858–63. https://doi.org/10.1073/pnas.1209751109.
[23] Melnikov A, Rogov P, Wang L, Gnirke A, Mikkelsen TS. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. e112–e112 Nucleic Acids Res 2014;42. https://doi.org/10.1093/nar/gku511.
[24] Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J. Massively parallel single-amino-acid mutagenesis. Nat Methods 2015;12:203–6. https://doi.org/10.1038/nmeth.3223.
[25] McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature 2012;491:138–42. https://doi.org/10.1038/nature11500.
[26] Melamed D, Young DL, Gamble CE, Miller CR, Fields S. Deep mutational scanning of an RRM domain of the saccharomyces cerevisiae poly (a)-binding protein. Rna 2013;19:1537–51.
[27] Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J Mol Biol 2013;425:1363–77. https://doi.org/10.1016/j.jmb.2013.01.032.
[28] Roscoe BP, Bolon DNA. Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast. J Mol Biol 2014;426:2854–70. https://doi.org/10.1016/j.jmb.2014.05.019.
[29] Mavor D, Barlow K, Thompson S, Barad BA, Bonny AR, Cario CL, et al. Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. eLife 2016;5:e15802. https://doi.org/10.7554/eLife.15802.
[30] Romero PA, Tran TM, Abate AR. Dissecting enzyme function with microfluidic-based deep mutational scanning. Proc Natl Acad Sci 2015;112:7159–64. https://doi.org/10.1073/pnas.1422285112.
[31] Li C, Qian W, Maclean CJ, Zhang J. The fitness landscape of a tRNA gene. Science 2016;352:837–40. https://doi.org/10.1126/science.aae0568.
[32] Julien P, Miñana B, Baeza-Centurion P, Valcárcel J, Lehner B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. Nat Commun 2016;7:11558. https://doi.org/10.1038/ncomms11558.
[33] Aakre CD, Herrou J, Phung TN, Perchuk BS, Crosson S, Laub MT. Evolving new protein-protein interaction specificity through promiscuous intermediates. Cell 2015;163:594–606. https://doi.org/10.1016/j.cell.2015.09.055.
[34] Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, et al. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. Proc Natl Acad Sci 2013;110:E1263–72. https://doi.org/10.1073/pnas.1303309110.
[35] Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M. Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. Mol Biol Evol 2016;33:268–80. https://doi.org/10.1093/molbev/msv211.
[36] Lapedes, A.; Giraud, B.; Jarzynski, C. Using sequence alignments to predict protein structure and stability with high accuracy. ArXiv Prepr. ArXiv12072484; 2012.
[37] Mann JK, Barton JP, Ferguson AL, Omarjee S, Walker BD, Chakraborty A, Ndung'u T. The fitness landscape of HIV-1 Gag: advanced modeling approaches and validation of model predictions by in vitro testing. PLOS Comput Biol 2014;10:e1003776. https://doi.org/10.1371/journal.pcbi.1003776.
[38] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9. https://doi.org/10.1038/nmeth0410-248.
[39] Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4. https://doi.org/10.1093/nar/gkg509.
[40] Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. BMC Genom 2015;16:S1. https://doi.org/10.1186/1471-2164-16-S8-S1.
[41] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5. https://doi.org/10.1038/ng.2892.
[42] Derbel H, Giacoletto CJ, Benjamin R, Chen G, Schiller MR, Liu Q. Accurate prediction of transcriptional activity of single missense variants in HIV tat with deep learning. Int J Mol Sci 2023;24:6138. https://doi.org/10.3390/ijms24076138.
[43] Derbel H, Zhao Z, Liu Q. Accurate prediction of functional effect of single amino acid variants with deep learning. Comput Struct Biotechnol J 2023;21:5776–84. https://doi.org/10.1016/j.csbj.2023.11.017.
[44] Hopf TA, Ingraham JB, Poelwijk FJ, Schärfe CPI, Springer M, Sander C, et al. Mutation effects predicted from sequence co-variation. Nat Biotechnol 2017;35:128–35. https://doi.org/10.1038/nbt.3769.

[45] Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature 2021;599:91–5. https://doi.org/10.1038/s41586-021-04043-8.
[46] Riesselman AJ, Ingraham JB, Marks DS. Deep generative models of genetic variation capture the effects of mutations. Nat Methods 2018;15:816–22. https://doi.org/10.1038/s41592-018-0138-4.
[47] Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci 2021;118:e2016239118.
[48] Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. Synth Biol 2021.
[49] Melamed D, Young DL, Gamble CE, Miller CR, Fields S. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. RNA 2013;19:1537–51. https://doi.org/10.1261/rna.040709.113.
[50] Pokusaeva VO, Usmanova DR, Putintseva EV, Espinar L, Sarkisyan KS, Mishin AS, et al. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLoS Genet 2019;15:e1008079. https://doi.org/10.1371/journal.pgen.1008079.
[51] Aakre CD, Herrou J, Phung TN, Perchuk BS, Crosson S, Laub MT. Evolving new protein-protein interaction specificity through promiscuous intermediates. Cell 2015;163:594–606. https://doi.org/10.1016/j.cell.2015.09.055.
[52] Ding D, Green AG, Wang B, Lite T-LV, Weinstein EN, Marks DS, et al. Coevolution of interacting proteins through non-contacting and non-specific mutations 2022.
[53] Staller MV, Holehouse AS, Swain-Lenz D, Das RK, Pappu RV, Cohen BA. A high-throughput mutational scan of an intrinsically disordered acidic transcriptional activation domain. Cell Syst 2018;6:444–455.e6. https://doi.org/10.1016/j.cels.2018.01.015.
[54] Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, et al. Local fitness landscape of the green fluorescent protein. Nature 2016;533:397–401. https://doi.org/10.1038/nature17995.
[55] Gonzalez Somermeyer L, Fleiss A, Mishin AS, Bozhanova NG, Igolkina AA, Meiler J, et al. Heterogeneity of the GFP fitness landscape and data-driven protein design. eLife 2022;11:e75842. https://doi.org/10.7554/eLife.75842.
[56] Chen Y, Hu R, Li K, Zhang Y, Fu L, Zhang J, et al. Deep mutational scanning of an oxygen-independent fluorescent protein CreiLOV for comprehensive profiling of mutational and epistatic effects. ACS Synth Biol 2023;12:1461–73. https://doi.org/10.1021/acssynbio.2c00662.
[57] Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A., Recursive deep models for semantic Compositionality Over a Sentiment Treebank.
[58] Hsu, C.; Verkuil, R.; Liu, J.; Lin, Z.; Hie, B.; Sercu, T., Learning inverse folding from millions of predicted structures; 2022.
[59] Manfredi M, Savojardo C, Martelli PL, Casadio R. E-pRSA: Embeddings improve the prediction of residue relative solvent accessibility in protein sequence. J Mol Biol 2024;436:168494. https://doi.org/10.1016/j.jmb.2024.168494.
[60] Kingma, D.P.; Ba, J. Adam: a method for stochastic optimization; 2017.
[61] Goller C, Kuchler A. Washington, DC, USA: IEEE; 1996. p. 347–52. Learning task-dependent distributed representations by backpropagation through structureIn. In: Proceedings of the international conference on neural networks (ICNN'96)1.
[62] Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes 2013.
[63] Luo Y, Jiang G, Yu T, Liu Y, Vo L, Ding H, et al. ECNet Is an evolutionary context-integrated deep learning framework for protein engineering. Nat Commun 2021;12:5743. https://doi.org/10.1038/s41467-021-25976-8.
[64] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Proc NAACL-HLT 2019:4171–86.
[65] Hopf TA, Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, et al. The EVcouplings Python framework for coevolutionary sequence analysis. Bioinformatics 2019;35:1582–4. https://doi.org/10.1093/bioinformatics/bty862.
[66] Notin, P.; Kollasch, A.W.; Ritter, D.; van Niekerk, L.; Paul, S.; Spinner, H., ; et al. ProteinGym: large-scale benchmarks for protein fitness prediction and design.
[67] Apweiler R. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2004;32:115D–119. https://doi.org/10.1093/nar/gkh131.
[68] Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res 2021;49:D412–9. https://doi.org/10.1093/nar/gkaa913.