Supplementary information

Supplemental Figure 1 | Design of the protein point cloud and the structure embedding module. For each type of protein point cloud, we trained a sub-network that only contains the structure embedding module and structure decoder module shown in Fig. 1. Each sub-network takes a specific type of point cloud as input to learn protein representations. The enzyme commission number annotation benchmark is used for performance evaluation (Methods). We compared the representations generated by the structure embedding module via an MLP classifier and reported the average precision (AP). **a,** Protein point clouds contain diverse point features. Compared with the naive unordered and homogeneous point cloud (Naive), the optimized protein point cloud (Optimized) is an ordered and heterogeneous set of alpha C atoms that belongs to each amino acid. The type of the residue and the connection between them are preserved. Protein point cloud without residue annotation was denoted as Masked. **b,** Type of atom used in the protein point cloud. Protein point cloud constructed by alpha C (CA) atoms, beta C (CB) atoms and N atoms were evaluated, respectively. **c,** Number of layers of the structure embedding module. The masked protein point cloud constructed by alpha C atoms was used during the evaluation.

Supplemental Figure 2 | Performance comparison on downstream tasks. a, We integrated our multimodal representations to CLEAN (denoted as Ours-CLEAN) and evaluated its performance on the New-392 database. **b,** Comparison of CLEAN-Ours, CLEAN, and other top-ranked models on the Price-149 database. **c-e,** Performance comparison for binary PPI prediction, including AUROC, AUPR, and F1 scores, on three virus-human PPI datasets, Denovo **(c)**, EBOLA **(d)**, and H1N1 **(e)**. **f-g,** Performance comparison for multi-class PPI prediction on the SHS148K **(f)** and STRING **(g)** dataset. All models are evaluated using two heuristic schemes: breadth-first search (BFS) and depth-first search (DFS).

Supplemental Figure 3 | Robustness tests on downstream tasks. a, Robustness test of our proposed model (denoted as Ours) and several important baselines under different sequence/structure similarity cutoffs between training and test sets. **b,** Performance of our model on test PPIs of three species (Mouse, Fly, and E.coli) in D-SCRIPT with regard to the sequence

identity to the training set. **c,** Performance comparison on the Fly PPIs of D-SCRIPT with regard to the sequence identity to the training set.

Supplemental Figure 4 | Extensive generalization tests on downstream tasks. a, A violin plot that shows the mean accuracy of our proposed model on 384 reaction classes with different numbers of proteins in the training set. **b,** Comparison of our proposed model (denoted as Ours) and several important baselines in one-shot reaction prediction. Each reaction class in the training set contains only one protein. **c,** Performance comparison for binary PPI prediction on the cross-species PPI dataset.

Supplemental Figure 5 | Extensive evaluations of the structure perception ability of our proposed model. a, Average performance comparison on a fold classification benchmark. **b,** Comparison of accuracy on 3-class (Q3) secondary structure classification benchmark. **c,** Three additional benchmarks related to local structure properties, which include 3-class secondary structure classification on the PDB-100 dataset, solvent accessible surface area (SASA) prediction, and b-factor prediction on the CATH-100 dataset. Specifically, SASA and b-factor are closely related to protein folding and stability. **d,** Performance of secondary structure classification on CLEAN dataset. **e,** An example of a secondary structure prediction map on real protein. Residues with a higher probability corresponding to β-sheet are highlighted in blue, and a higher probability corresponding to α-helix are highlighted in red. The predicted secondary structures are highly consistent with the ground truth.

Supplemental Figure 6 | Extensive evaluations of the functional site identification ability of our proposed model. a, Functional sites identification score on the CLEAN dataset. We compared our proposed model (denoted as Ours) with other leading methods. **b,** A box plot that shows the function prediction performance of multimodal representations under different percentage attention shuffling. **c,** An example of function sites identification on a real protein. More salient residues are highlighted in red.

Supplemental Figure 7 | Structure signatures of KKA2_KLEPN. a, Quantification of local structure perception ability of our multimodal deep representation learning model on

KKA2_KLEPN. **b,** Visualization of secondary structure prediction map on the real structure of KKA2_KLPEN. Residues with a higher probability corresponding to β-sheet are highlighted in blue, and a higher probability corresponding to α-helix are highlighted in red.

Supplemental Figure 8 | Performance of ProMEP in predicting mutational effects on DMS datasets with various mutational depths. ProMEP consistently shows superior performance for multi-mutations.

Supplemental Figure 9 | Visualization of 3 proteins annotated with functional sites. The active/binding sites of these proteins can be accurately identified by ProMEP, including Glu59 in P11436, Asp96 in P0DP23 and Ser108 in P37957.

Supplemental Figure 10 | Performance of ProMEP that uses different sources of protein structure data on other 17 DMS datasets. Structures predicted by Alphafold2 lead to slightly better performance in predicting mutational effects.

Supplemental Figure 11 | Quantification of multi-scale structure signatures on TnpB. a, Secondary structure signatures of TnpB can be accurately captured by our proposed model. **b,** Our proposed model identifies functional sites of TnpB.

Supplemental Figure 12 | Workflow of Indel characterization.

Supplemental Table 1. Performance comparison on function annotation datasets.

Supplemental Table 2. Performance comparison on PPI datasets.

Supplemental Table 3. Description of 20 DMS datasets.

Supplemental Table 4. TnpB variants with single mutation.

Supplemental Table 5. TnpB variants with multiple mutations.

Supplemental Table 6. Description of 15 function-related datasets

Supplemental Table 7. Number of proteins in virus-human datasets.

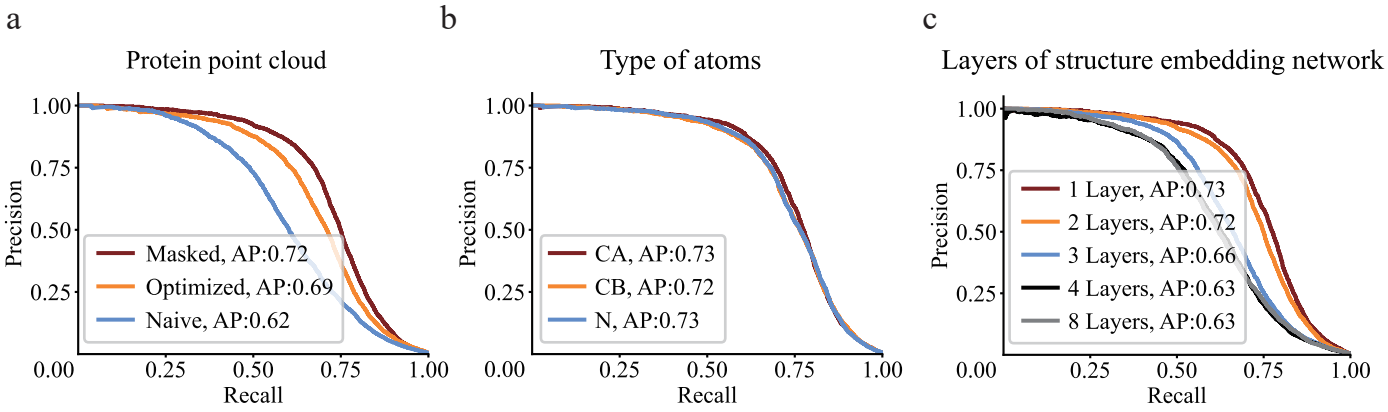
Supplemental Table 8. Number of proteins in the test set with 30%, 40%, 50%, 70% and 95% sequence identity to the training set.

Supplemental Table 9. Number of proteins in the test set with 30%, 40%, 50%, 70% and 95% structure identity to the training set.

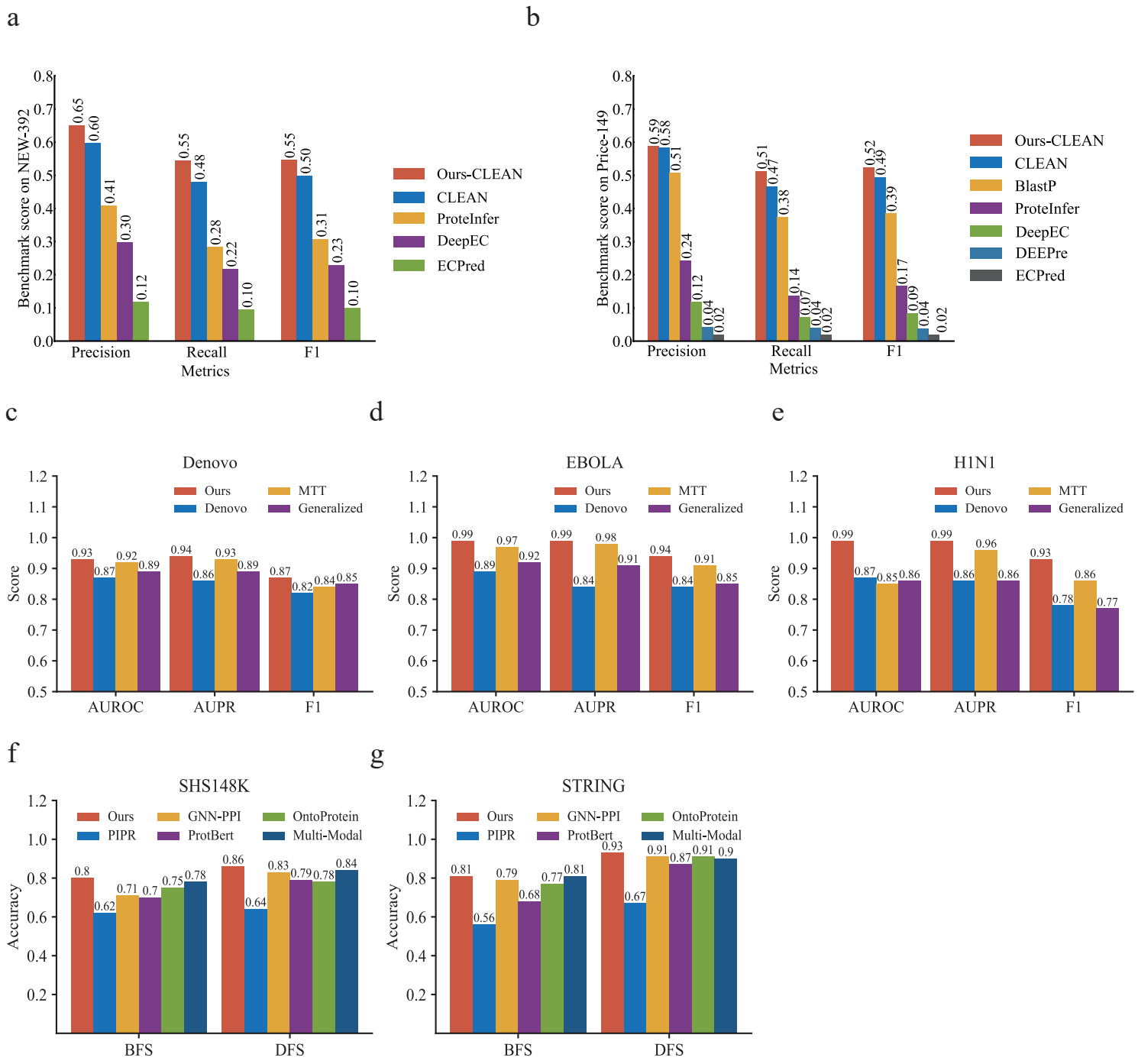
Supplemental Table 10. Fold class of 13,265 domains from the SCOPe database.

Supplemental Table 11. TnpB target site sequences in this study.

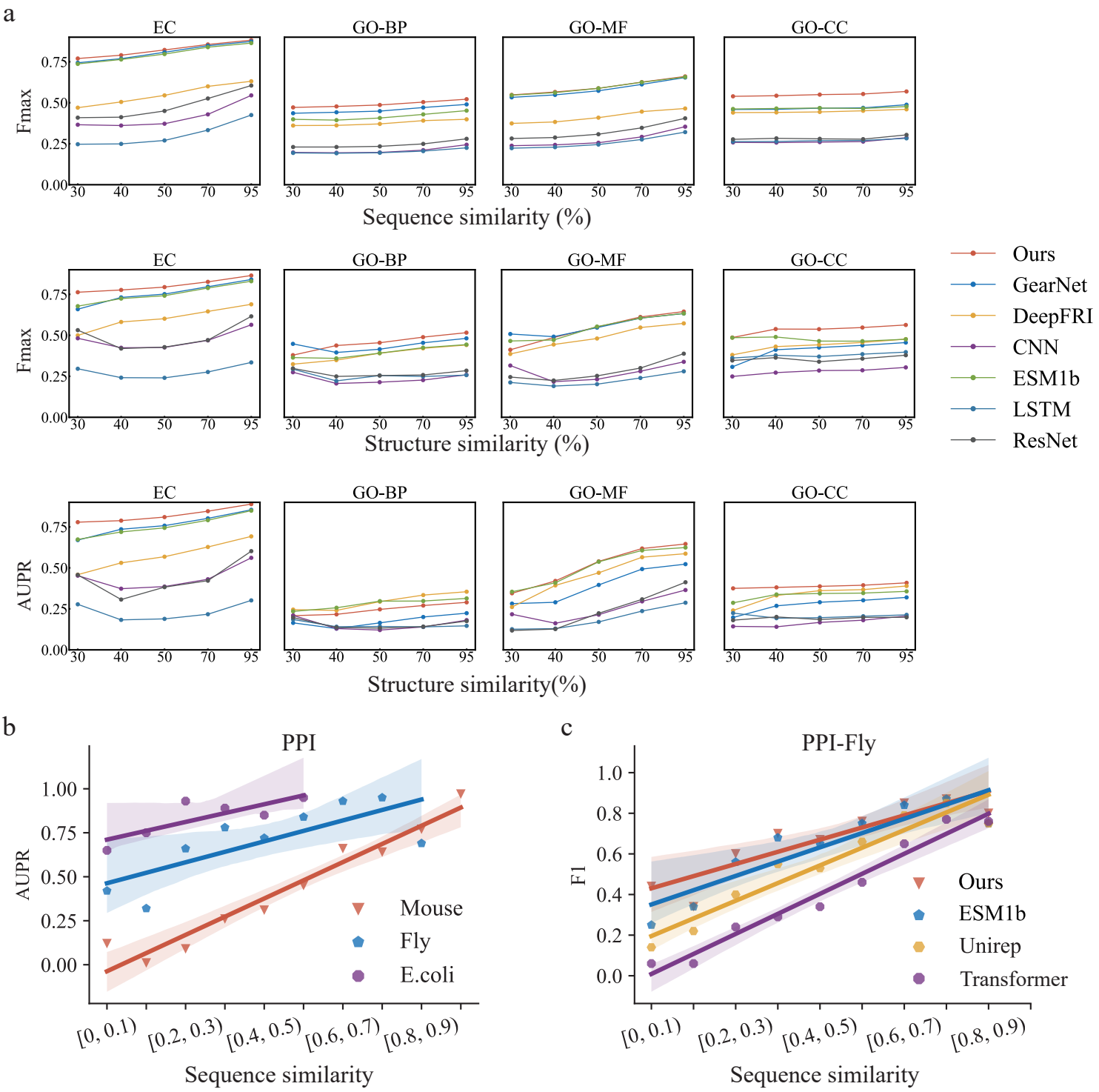
Supplemental Table 12. Sequences of PCR1 primers for Deep sequencing.



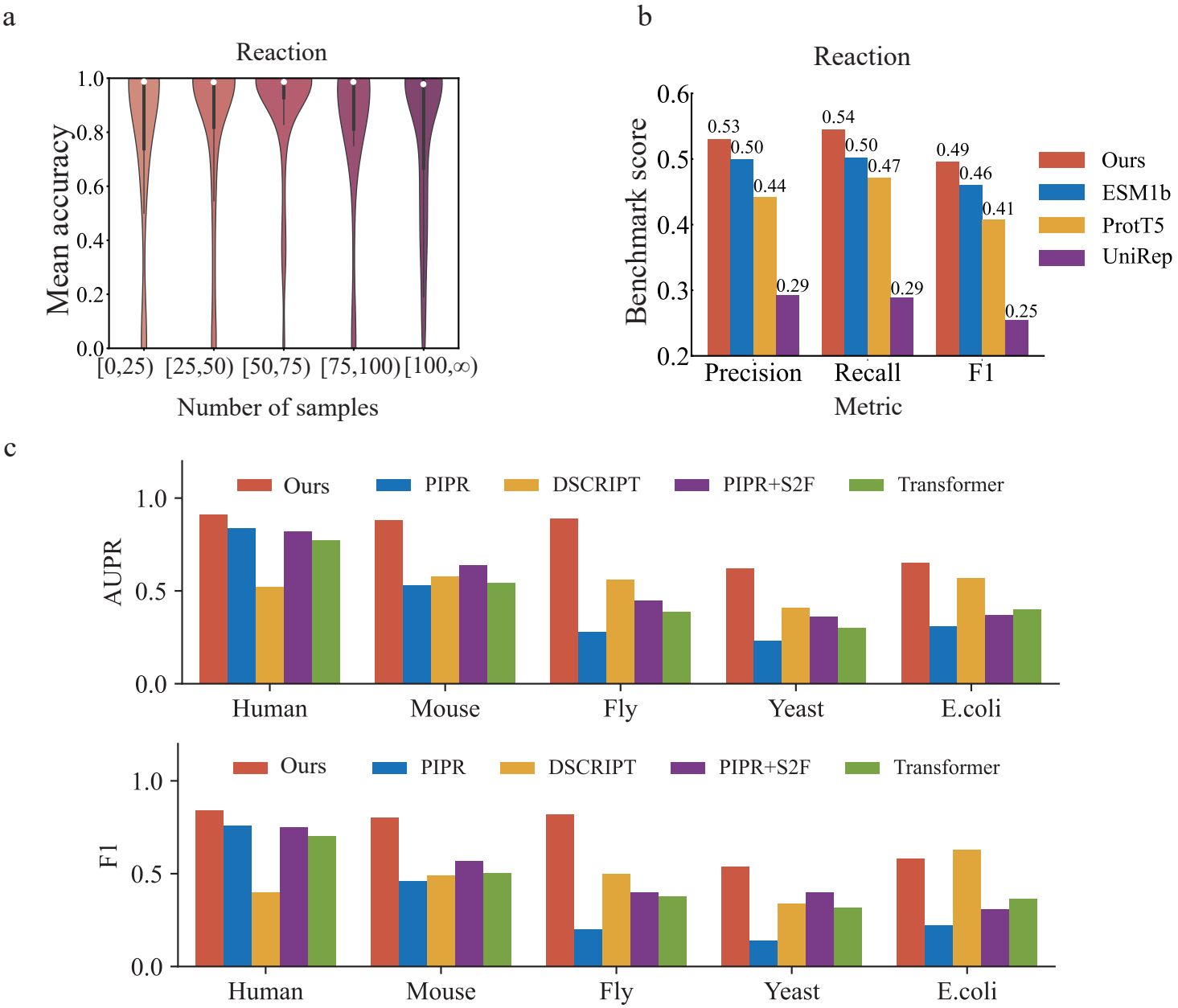
Supplemental Figure 1



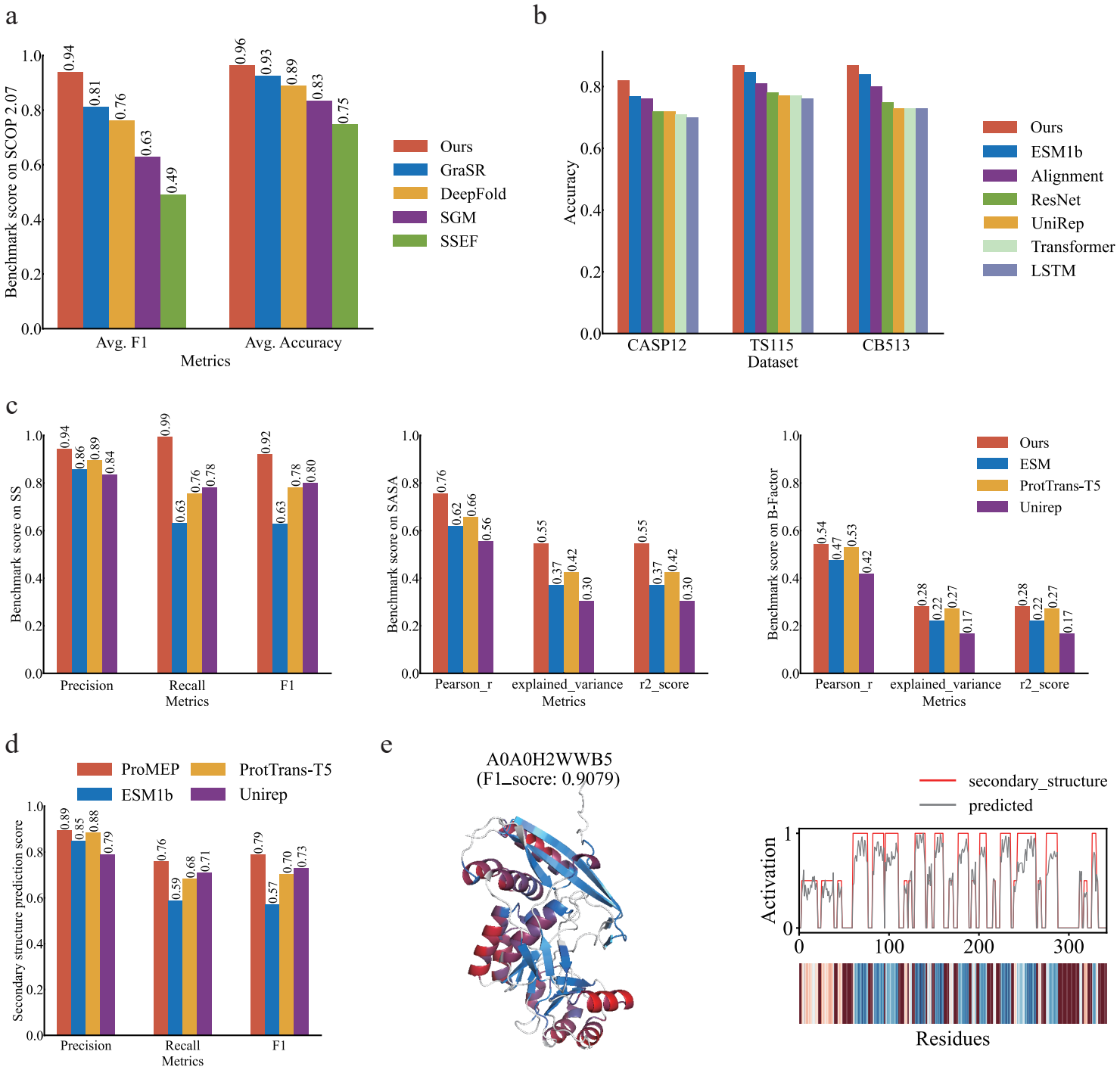
Supplemental Figure 2



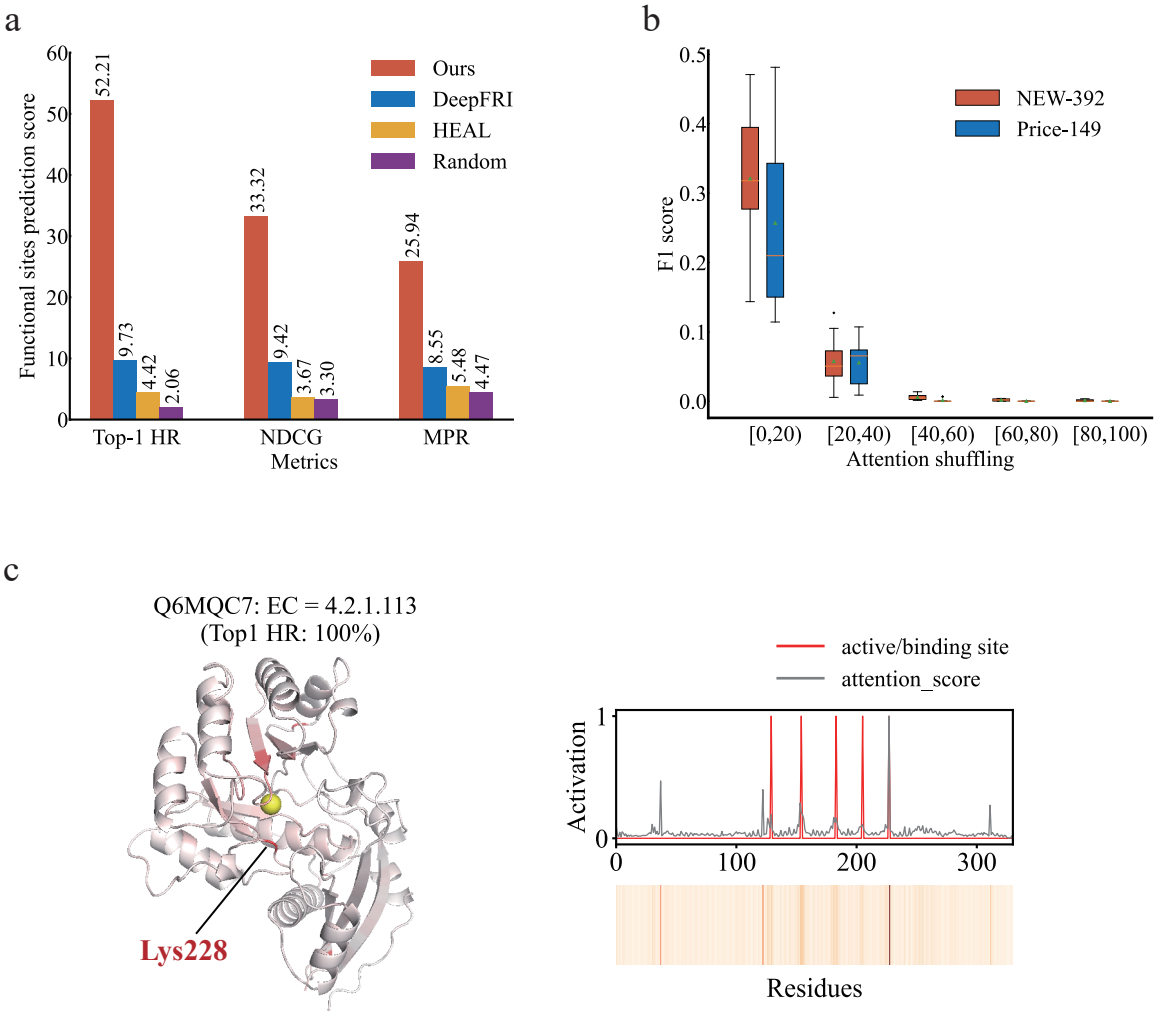
Supplemental Figure 3



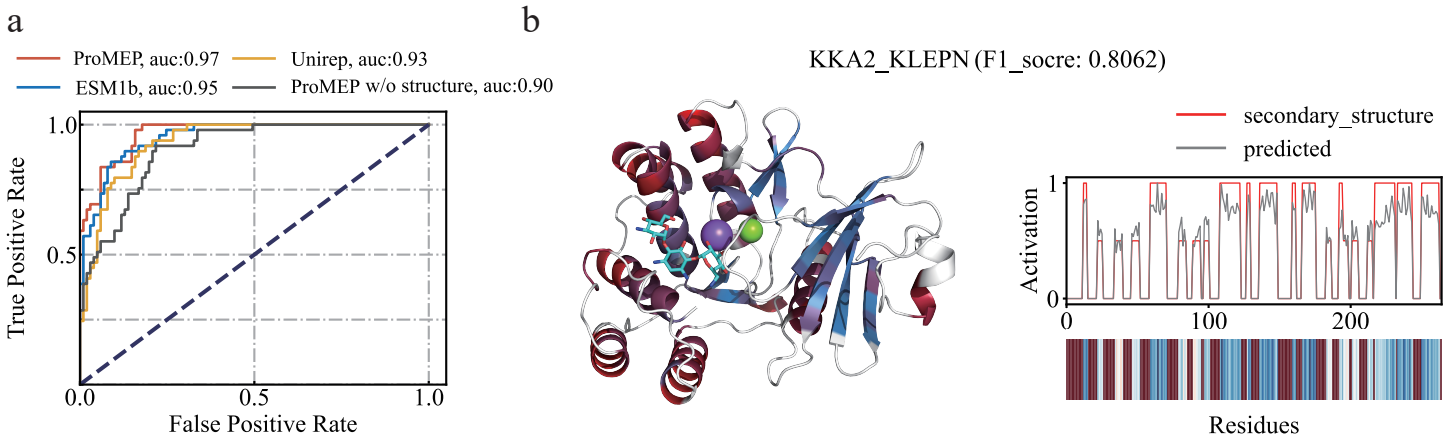
Supplemental Figure 4



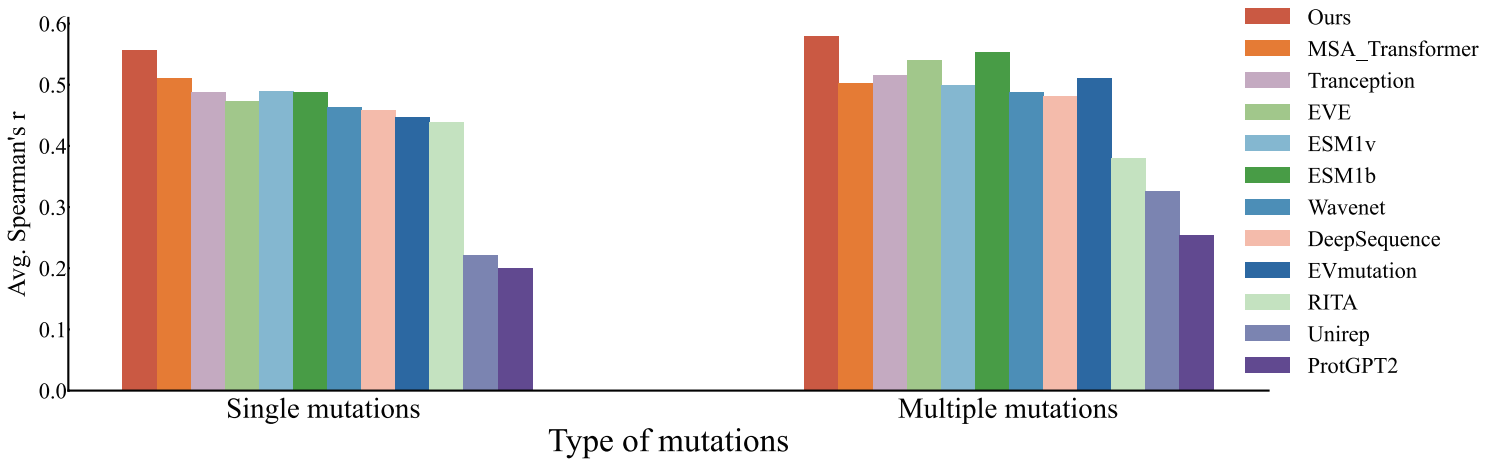
Supplemental Figure 5



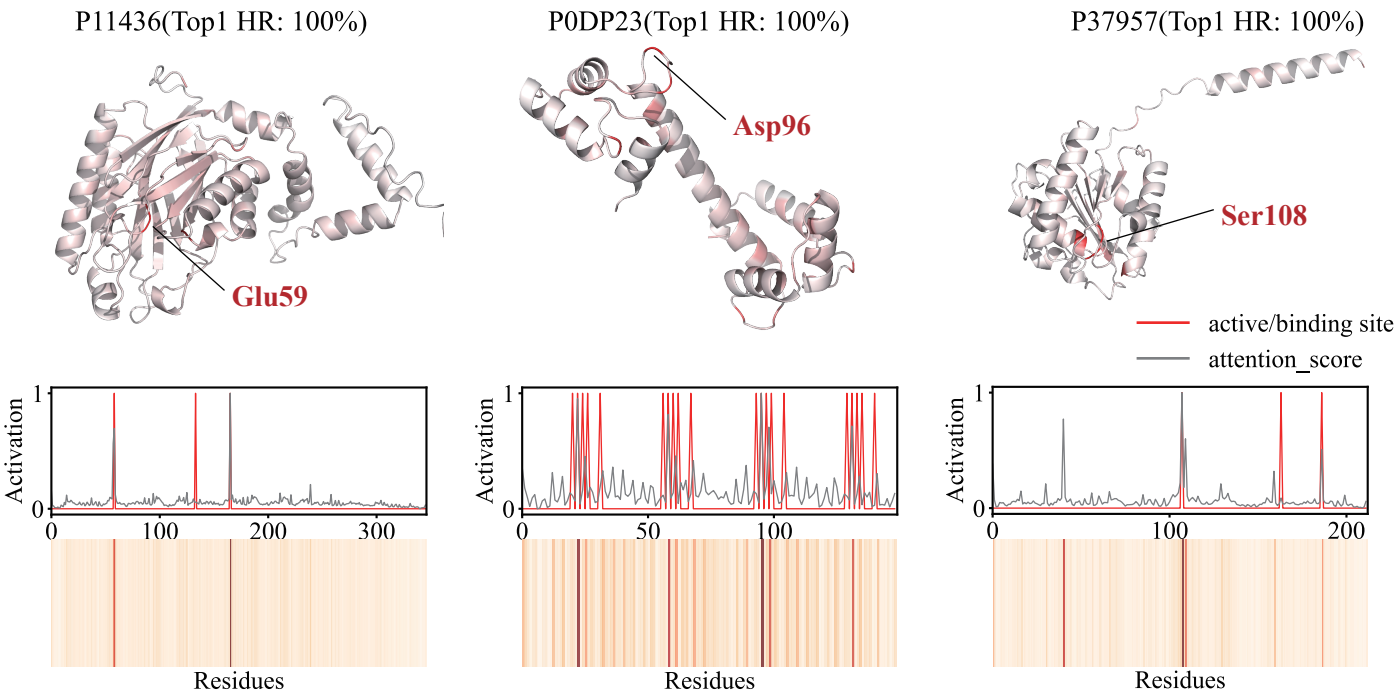
Supplemental Figure 6



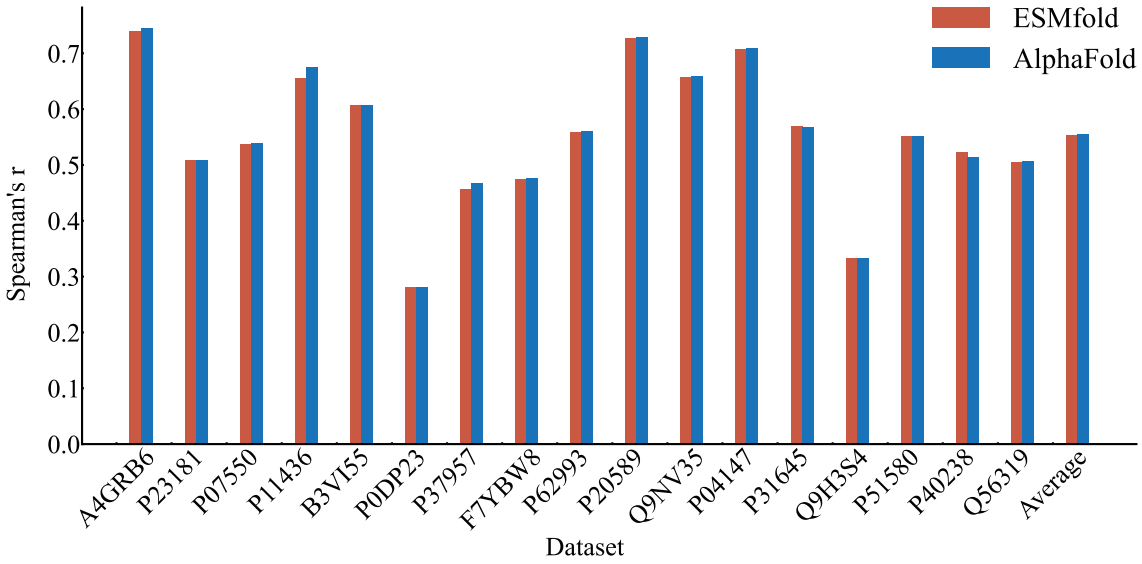
Supplemental Figure 7



Supplemental Figure 8

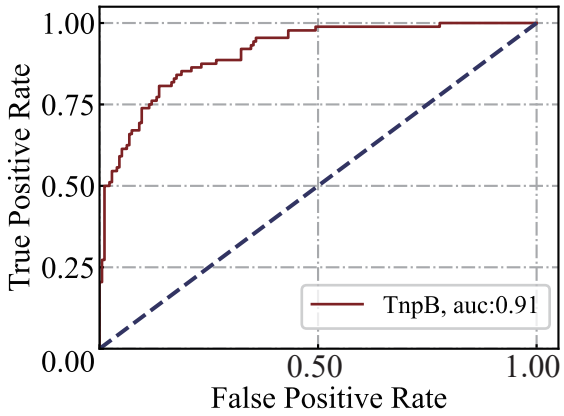


Supplemental Figure 9

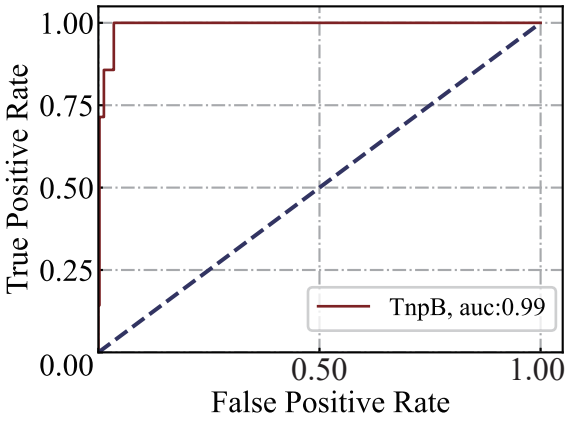


Supplemental Figure 10

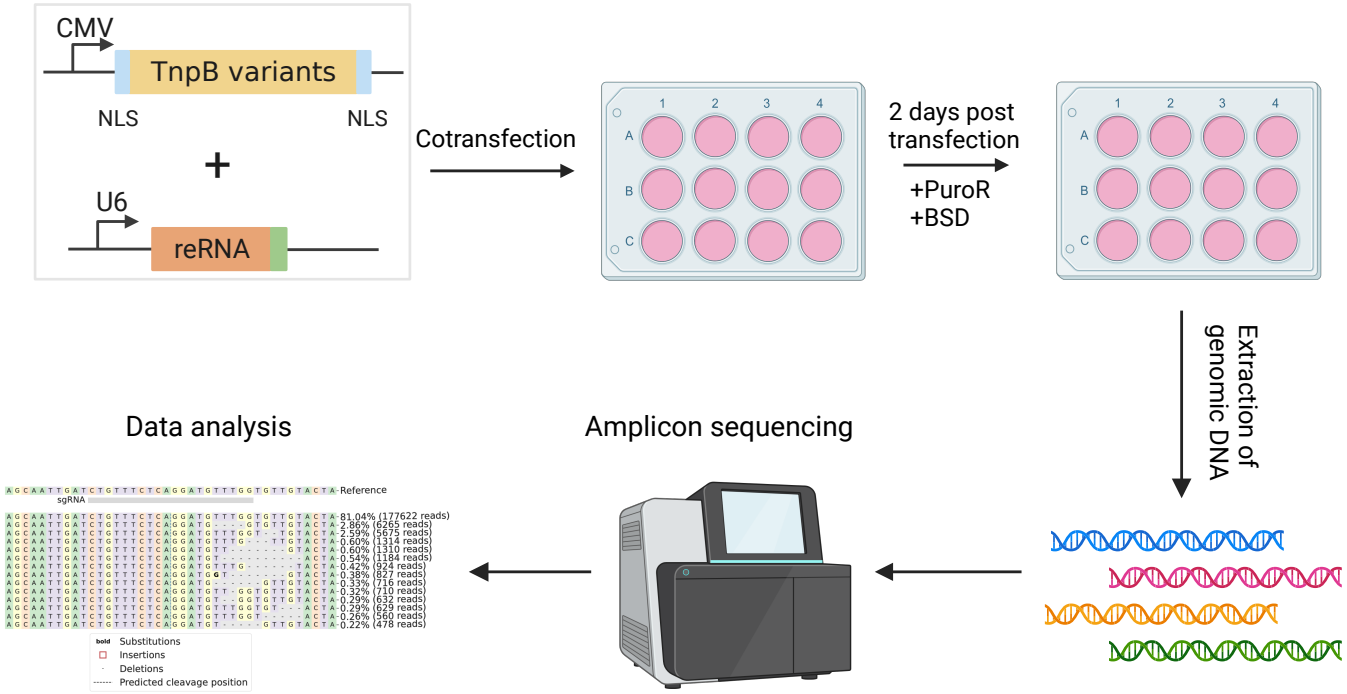
a



b



Supplemental Figure 11



Supplemental Figure 12

Type	Method	EC		GO-BP		GO-MF		GO-CC		Reaction
		Fmax	AUPR	Fmax	AUPR	Fmax	AUPR	Fmax	AUPR	Accuracy
w/o pretraining	CNN [75]	0.545	0.526	0.244	0.159	0.354	0.351	0.287	0.204	0.517
	ResNet [76]	0.605	0.590	0.280	0.205	0.405	0.434	0.304	0.214	0.241
	LSTM [76]	0.425	0.414	0.225	0.156	0.321	0.334	0.283	0.192	0.110
	Transformer [76]	0.238	0.218	0.264	0.156	0.211	0.177	0.405	0.210	0.266
	GCN [78]	0.320	0.319	0.252	0.136	0.195	0.147	0.329	0.175	0.673
	GAT [79]	0.368	0.320	0.284	0.171	0.317	0.329	0.385	0.249	0.556
	GVP [80]	0.489	0.482	0.326	0.224	0.426	0.458	0.420	0.279	0.655
	GraphQA [81]	0.509	0.543	0.308	0.199	0.329	0.347	0.413	0.265	0.608
Sequence pretraining	UniRep [26]	0.698	0.554	0.375	0.276	0.505	0.523	0.407	0.3	0.733
	ESM [27]	0.864	0.889	0.452	**0.332**	**0.657**	0.639	0.477	0.324	0.831
	ProtTrans [28]	0.838	0.859	0.279	0.188	0.456	0.464	0.408	0.234	0.72.2
Structure pretraining	GearNet [33]	**0.874**	0.892	**0.490**	0.292	0.654	0.596	0.488	0.336	**0.875**
Sequence+Structure pretraining	DeepFRI [43]	0.631	0.547	0.399	0.282	0.465	0.462	0.460	0.363	0.633
	LM-GVP [77]	0.664	0.710	0.417	0.302	0.515	0.580	**0.527**	**0.423**	-
	Ours	**0.882**	**0.908**	**0.521**	0.299	**0.660**	**0.656**	**0.568**	0.410	**0.901**

Supplemental Table 1

Type	Method	PPI-Mouse		PPI-Fly		PPI-Ecoli	
		F1	AUPR	F1	AUPR	F1	AUPR
w/o pretraining	CNN [75]	0.418	0.526	0.294	0.317	0.435	0.457
	ResNet [76]	0.361	0.515	0.317	0.434	0.473	0.507
	LSTM [76]	0.467	0.549	0.417	0.431	0.787	0.904
	Transformer [76]	0.503	0.543	0.377	0.388	0.365	0.401
Sequence pretraining	UniRep [26]	0.498	0.566	0.583	0.620	0.514	0.584
	ESM [27]	0.580	**0.686**	0.692	0.764	**0.597**	**0.695**
	ProtTrans [28]	**0.605**	0.685	**0.711**	**0.786**	0.590	0.679
Sequence+Structure pretraining	Ours	**0.620**	**0.707**	**0.720**	**0.790**	**0.629**	0.667

Supplemental Table 2

UniProt ID	Length (aa)	Source organism	Measurements	Includes Multiple mutants
A4GRB6	266	Pseudomonas aeruginosa	Antibiotics resistance	False
P23181	177	Pseudomonas aeruginosa	Antibiotics resistance	False
P07550	413	Homo sapiens	Receptor activity	False
P11436	346	Pseudomonas aeruginosa	Growth	False
B3VI55	439	Lipomyces starkeyi (Oleaginous yeast)	Growth	False
P0DP23	149	Homo sapiens	complementation	False
P62554	101	Escherichia coli	Growth	False
P37957	212	Bacillus subtilis	thermostability	False
F7YBW8	93	Mesorhizobium opportunistum (strain LMG 24607 / HAMBI 3007 / WSM2075)	Growth (antitoxin neutralization of ParE3)	True
P62993	217	Homo sapiens	Growth	True
P20589	330	Haemophilus aegyptius	Activity	False
Q9NV35	164	Homo sapiens	VAMP-seq, drug sensitivity	False
P04637	393	Homo sapiens	Growth	False
P84126	254	Thermus thermophilus	Growth	False
P04147	577	Saccharomyces cerevisiae S288C	Growth	True
P31645	630	Homo sapiens	Fluorescence	False
Q9H3S4	243	Homo sapiens	complementation	False
P51580	245	Homo sapiens	Protein stability	False
P40238	635	Homo sapiens	Growth	False
Q56319	252	Thermus thermophilus	Growth	False

Supplemental Table 3

Variant	Mutation	Average fold change	P-value
V1.1	S57R	1.38838	0.001152
V1.2	S217R	1.015612	0.70847
V1.3	S72R	1.557904	0.000395
V1.4	L61R	1.444216	0.00104
V1.5	Y388R	0.87958	0.020112
V1.6	L406R	1.076939	0.168706
V1.7	K44R	1.073025	0.164516
V1.8	H403R	1.096893	0.553405
V1.9	L398R	0.219713	0.000003
V1.10	T405R	0.92828	0.025489
V1.11	A198R	0.012417	0.000001
V1.12	V171R	0.217586	0.00002
V1.13	A298R	0.014471	0.000001
V1.14	A78R	0.010568	0.000001
V1.15	F93R	0.373669	0.000028
V1.16	L136R	0.232685	0.000115
V1.17	D297R	0.427312	0.000014
V1.18	I127R	0.198395	0.000212
V1.19	A89R	0.265094	0.000086
V1.20	I275R	0.009649	<0.000001
V1.21	L11R	0.116725	0.000013
V1.22	K112R	0.952634	0.39121
V1.23	Q128R	1.076945	0.025073
V1.24	I194R	0.021172	0.000001
V1.25	L398R	0.219703	0.000003
V1.26	K103R	1.263591	0.000548
V1.27	L180R	0.140701	0.000012
V1.28	T201R	0.008531	<0.000001
V1.29	Y239R	0.265438	0.00003
V1.30	L136R	0.232685	0.000115

Supplemental Table 4

Space	Variant	Mutation	Average fold change	P-value
Double	V2.1	S72R-S57R	1.760274	0.000425
	V2.2	L61R-S57R	1.381481	0.001966
	V2.3	S72R-L61R	1.7602	0.000221
	V2.4	S69R-S57R	1.620446	0.000209
	V2.5	K145R-S57R	1.837732	0.000139
	V2.6	S57R-K207R	1.804682	0.000082
	V2.7	S57R-P391R	1.730274	0.000143
	V2.8	L382R-S57R	1.597831	0.000282
	V2.9	S267R-S57R	1.737728	0.001
	V2.10	Q258R-S57R	0.318336	0.000176
S72R-D	V2.11	S72R-S57R	1.760274	0.000425
	V2.12	S72R-L61R	1.7602	0.000221
	V2.13	S72R-S69R	1.760235	0.000224
	V2.14	S72R-K145R	1.496949	0.001116
	V2.15	S72R-K207R	1.279251	0.006511
	V2.16	S72R-P391R	1.461532	0.001813
	V2.17	S72R-L382R	1.446663	0.001261
	V2.18	S72R-S267R	1.339627	0.007234
	V2.19	S72R-Q258R	2.050806	0.000171
	V2.20	S72R-K337R	1.17935	0.032345
Triple	V3.1	S57R-L61R-S72R	1.563663	0.001376
	V3.2	S57R-S69R-S72R	1.862519	0.000072
	V3.3	K145R-S72R-S57R	1.843975	0.000191
	V3.4	S57R-S72R-K207R	1.861322	0.000138
	V3.5	L61R-S69R-S57R	1.706469	0.000589
	V3.6	L61R-K145R-S57R	1.89755	0.000519
	V3.7	P391R-S72R-S57R	1.908791	0.000141
	V3.8	S72R-L382R-S57R	1.855532	0.000326
	V3.9	K207R-L61R-S57R	1.58635	0.006778
	V3.10	S72R-S267R-S57R	2.232219	0.000093
S72R-T	V3.11	S72R-S57R-L61R	1.563663	0.001376
	V3.12	S72R-S57R-S69R	1.999542	0.000047
	V3.13	S72R-K145R-S57R	1.843975	0.000191
	V3.14	S72R-S57R-K207R	2.160273	0.000234
	V3.15	S72R-P391R-S57R	2.225668	0.00032
	V3.16	S72R-S57R-L382R	1.91437	0.000233
	V3.17	S72R-S267R-S57R	1.626623	0.000727
	V3.18	S72R-Q258R-S57R	0.747634	0.020375
	V3.19	S72R-S57R-K337R	1.923069	0.010991
	V3.20	S72R-Q99R-S57R	2.389401	0.000104

Supplemental Table 5

Task	Dataset	Train	Valid	Test	Number of classes	Classifier	Size of representation of each protein
EC	PDB	15,550	1,729	1,919	538	MLP	1,280
	New-392	227,362	-	392	5,242	CLEAN	1,280
	Price-149	227.362	-	149	5,242	CLEAN	1,280
	Reaction	29,215	2,562	5,651	384	MLP	1,280
GO	BP	29,898	3,322	3,415	1,943	MLP	1,280
	CC	29,898	3,322	3,415	320	MLP	1,280
	MF	29,898	3,322	3,415	489	MLP	1,280
PPI (Cross-species)	Fly	-	-	55,000	2	MLP	1,280
	Mouse	-	-	55,000	2	MLP	1,280
	Ecoli	-	-	22,000	2	MLP	1,280
	Human	421,746	52,725	-	2	MLP	1280
PPI (Virus-human)	Denovo	9,754	-	850	2	MLP	1,280
	EBOLA	22,682	-	300	2	MLP	1,280
	H1N1	21,616	-	762	2	MLP	1,280
PPI (Multi-class)	SHS148K	35,585	-	8,903	7	GNN	L * 1280
	STRING	473,112	-	120,287	7	GNN	L * 1280

Supplemental Table 6

Dataset	Human proteins	Virus proteins
Denovo	2,341	445
EBOLA	7,816	659
H1N1	6,636	641

Supplemental Table 7

Dataset	30%	40%	50%	70%	95%
EC-PDB	720	902	1,117	1,476	1,919
GO-MF GO-BP GO-CC	1,717	1,937	2,199	2,733	3,416

Supplemental Table 8

Dataset	30%	40%	50%	70%	95%
EC-PDB	117	369	647	1056	1539
GO-MF GO-BP GO-CC	30	185	738	2,045	2,933

Supplemental Table 9

Fold	Number of proteins
a : All alpha proteins	2286
b : All beta proteins)	2757
c : Alpha and beta proteins (a/b)	4148
d : Alpha and beta proteins (a+b)	3378
e : Multi-domain proteins (alpha and beta)	279
f : Membrane and cell surface proteins and peptides	213
g : Small proteins	204

Supplemental Table 10

EMX1 site1	CTGTTTCTCAGGATGTTTGG
AGBL1	TGTTGGCTCAAACACCAGAT
IFNG	GACCTTCTTTGCTCCAAAAC
CLIC4	CTATCTAAAATAAAATGGAT
EMX1 site2	TCACTTACATAGATGTTTCC
RNF2 site1	CCTTGGATTTATAAAGTATT
RNF2 site2	GAACATTTCTCTACCCAGGA

Supplemental Table 11

EMX1-site1-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGggtggttcaggcctccttcccac
EMX1-site1-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAggtggttcaggcctccttcccac
EMX1-site1-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAggtggttcaggcctccttcccac
EMX1-site1-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGggtggttcaggcctccttcccac
EMX1-site1-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCggtggttcaggcctccttcccac
EMX1-site1-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCcaagatgctaagtgatgacagg
EMX1-site1-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGcaagatgctaagtgatgacagg
EMX1-site1-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTcaagatgctaagtgatgacagg
EMX1-site1-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAcaagatgctaagtgatgacagg
EMX1-site1-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAcaagatgctaagtgatgacagg
AGBL1-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGgcctctgatactgacccatcagg
AGBL1-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAgcctctgatactgacccatcagg
AGBL1-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAgcctctgatactgacccatcagg
AGBL1-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGgcctctgatactgacccatcagg
AGBL1-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCgcctctgatactgacccatcagg
AGBL1-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCgttgcgagcgactgtatacata
AGBL1-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGgttgcgagcgactgtatacata
AGBL1-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTgttgcgagcgactgtatacata
AGBL1-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAgttgcgagcgactgtatacata
AGBL1-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAgttgcgagcgactgtatacata
IFNG-site2-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGgctgagaagatgtgtgttctcct
IFNG-site2-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAgctgagaagatgtgtgttctcct
IFNG-site2-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAgctgagaagatgtgtgttctcct
IFNG-site2-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGgctgagaagatgtgtgttctcct
IFNG-site2-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCgctgagaagatgtgtgttctcct
IFNG-site2-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCgttggctagagacttgcagtgg
IFNG-site2-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGgttggctagagacttgcagtgg
IFNG-site2-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTgttggctagagacttgcagtgg
IFNG-site2-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAgttggctagagacttgcagtgg
IFNG-site2-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAgttggctagagacttgcagtgg
CLIC4-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGggcaggaggatcacctgaacctt
CLIC4-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAggcaggaggatcacctgaacctt
CLIC4-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAggcaggaggatcacctgaacctt
CLIC4-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGggcaggaggatcacctgaacctt
CLIC4-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCggcaggaggatcacctgaacctt
CLIC4-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCgaaccaccacatctggcccta
CLIC4-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGgaaccaccacatctggcccta
CLIC4-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTgaaccaccacatctggcccta
CLIC4-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAgaaccaccacatctggcccta
CLIC4-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAgaaccaccacatctggcccta
EMX1-site2-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGggcttctttaggtaaccaaagcc
EMX1-site2-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAggcttctttaggtaaccaaagcc
EMX1-site2-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAggcttctttaggtaaccaaagcc
EMX1-site2-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGggcttctttaggtaaccaaagcc
EMX1-site2-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCggcttctttaggtaaccaaagcc

EMX1-site2-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCcagagatggctactggttggac
EMX1-site2-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGcagagatggctactggttggac
EMX1-site2-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTcagagatggctactggttggac
EMX1-site2-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAcagagatggctactggttggac
EMX1-site2-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAcagagatggctactggttggac
RNF2-site1-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGgtcctgctggtcagaatcgttta
RNF2-site1-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAgtcctgctggtcagaatcgttta
RNF2-site1-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAgtcctgctggtcagaatcgttta
RNF2-site1-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGgtcctgctggtcagaatcgttta
RNF2-site1-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCgtcctgctggtcagaatcgttta
RNF2-site1-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCggcacttcaaagaaacctgaca
RNF2-site1-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGggcacttcaaagaaacctgaca
RNF2-site1-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTggcacttcaaagaaacctgaca
RNF2-site1-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAggcacttcaaagaaacctgaca
RNF2-site1-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAggcacttcaaagaaacctgaca
RNF2-site2-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGggccttcccagaaaagtagaact
RNF2-site2-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAggccttcccagaaaagtagaact
RNF2-site2-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAggccttcccagaaaagtagaact
RNF2-site2-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGggccttcccagaaaagtagaact
RNF2-site2-F5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCggccttcccagaaaagtagaact
RNF2-site2-R1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCgtccagggcctcatatgccaca
RNF2-site2-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGgtccagggcctcatatgccaca
RNF2-site2-R3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTgtccagggcctcatatgccaca
RNF2-site2-R4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAgtccagggcctcatatgccaca
RNF2-site2-R5	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTAgtccagggcctcatatgccaca

Supplemental Table 12