Supplementary Tables

Table S1. Input data type and featurization approaches in selected classical protein representations used in protein function prediction methods.

Input Data Type	Method	Utilized Feature
Protein Sequence	Cozzetto *et al.*[159]	Residue composition, Sequence length
	Lan *et al.*[160]	Identity based similarity score
	Hawkins *et al.*[161]	Identity based similarity score
	Cao and Cheng[162]	Identity based similarity score, E-value based probabilistic confidence score
	Piovesan *et al.*[163]	Identity-based similarity score
Homology	BLAST[54]	Identity based similarity score
	HMMER[55]	Identity based similarity score
	PFAM[57]	Functional/structural regions
	K-Sep[6]	Transition prob. between AA
Rule-based	Ensembl-Orthology[52]	Annotation transfer between orthologs
	UniRule2GO[50]	Expert Curation
	InterPro2GO[51]	Expert Curation
Composition	AAC[60]	AA Composition
	APAAC[62]	AA Composition + Physicochemical properties
Disordered regions	Cozzetto *et al.*[159]	Number of disordered regions
	Oates *et al.*[164]	One-hot encoding of disordered regions
Protein-protein interaction	Lan *et al.*[160]	Similarity score based on binary interaction vectors.
	Youngs *et al.*[165]	Interaction weighting matrices
	Piovesan *et al.*[163]	Binary interaction vectors
Gene expression	Lan *et al.*[160]	Expression based similarity score
Functional annotations	Sasidharan *et al.*[166]	Domain annotation list
	Cozzetto *et al.*[159]	Number of transmembrane residues, low complexity regions, glycosylated residues; localisation existence, Signal Peptide length and score
Literature texts	Van Landeghem *et al.*[167]	Named entity tags
Secondary structures	Cozzetto *et al.*[159]	Number of helices, sheets, coils
Physicochemical properties	Cozzetto *et al.*[159]	Molecular weight, Average hydrophobicity, Charge, Molar extinction coefficient, Isoelectric point, Aliphatic index
	APAAC[62]	Hydrophobicity + AA Composition

Table S2. The number of proteins for each GO category in the dataset used in the protein function prediction benchmark.

			Grouping GO terms in terms of the number of annotations		
			High	**Middle**	**Low**
Grouping GO terms considering the specificity of the term (i.e., location on the hierarchy of GO DAG)	**Shallow**	**BP**	3,386	896	27
		MF	3,039	613	41
		CC	7,186	413	55
	Normal	**BP**	3,288	452	28
		MF	1,324	369	31
		CC	6,478	562	47
	Specific	**BP**	2,025	623	42
		MF	0	204	41
		CC	0	413	54

Table S3. Identifiers of GO terms that are incorporated into each of the 25 different multi-task models constructed for the protein function prediction benchmark.

			Grouping GO terms in terms of the number of annotations		
			High	**Middle**	**Low**
Grouping GO terms considering the specificity of the term (i.e., location on the hierarchy of GO DAG)	**Shallow**	**BP**	GO:0007399 GO:0006259 GO:0007167 GO:0006886 GO:0051707	GO:0043488 GO:0043312 GO:0038096 GO:0006637 GO:0051091	GO:0032933 GO:0006348 GO:0042989 GO:0000054 GO:0039529
		MF	GO:0016818 GO:0022890 GO:0046872 GO:0004672 GO:0000981	GO:0036459 GO:0046943 GO:0005524 GO:0005244 GO:0022835	GO:0015187 GO:0008568 GO:0008508 GO:0005328 GO:0043138
		CC	GO:0005789 GO:1990234 GO:1903561 GO:0005740 GO:0070013	GO:0008076 GO:0005762 GO:0005747 GO:0036464 GO:0044853	GO:0000124 GO:0097165 GO:0008024 GO:0032590 GO:0033162
	Normal	**BP**	GO:0070647 GO:0016071 GO:0015833 GO:0048699 GO:0019752	GO:0071427 GO:0016573 GO:0006613 GO:0031146 GO:0051092	GO:0098703 GO:0061740 GO:0014808 GO:0000056 GO:0050482
		MF	GO:0016462 GO:0044212	GO:0003774 GO:0008227 GO:0004866 GO:0004714 GO:0004843	GO:0016934 GO:0022848 GO:0005314 GO:0003689 GO:0005335
		CC	GO:0070062 GO:0000228 GO:0031981 GO:0015630 GO:0005768	GO:0000502 GO:0005925 GO:0022627 GO:0030665 GO:0034705	GO:0036020 GO:0070081 GO:0044754 GO:1990454 GO:0031089
	Specific	**BP**	GO:0045944 GO:0001934 GO:1903507	GO:1903169 GO:0050773 GO:0071805 GO:0031124 GO:0000209	GO:0043984 GO:0034625 GO:0006231 GO:0046323 GO:0072539
		MF	-	GO:0061733 GO:0003777 GO:0004386	GO:0004957 GO:0008511 GO:0004571 GO:0017049 GO:0097199
		CC	-	GO:0005766 GO:0016591 GO:0101002 GO:0008021	GO:0019908 GO:0005767 GO:0032009 GO:0005736 GO:0099061

Table S4. Overall performance results in ontology-based protein function prediction benchmark for GO: **(a)** molecular function, **(b)** biological process, **(c)** cellular component.

(a)

Methods	**Accuracy**	**F1-Weighted**	**Precision**	**Recall**	**Hamming**
InterPro2GO	0.28	0.37	0.86	0.28	0.20
UniRule2GO	0.01	0.01	0.54	0.01	0.27
Ensembl-Orthology	0.11	0.20	0.89	0.12	0.24
BLAST	0.84	0.87	0.92	0.86	0.05
HMMER	**0.86**	0.89	0.91	**0.89**	0.05
K-Sep	0.71	0.81	0.85	0.81	0.09
APAAC	0.47	0.58	0.67	0.60	0.17
PFAM	0.81	0.86	**0.93**	0.84	0.06
AAC	0.39	0.41	0.49	0.44	0.20
ProtVec	0.52	0.64	0.68	0.67	0.17
Gene2Vec	0.44	0.53	0.59	0.53	0.20
Learned-Vec	0.59	0.68	0.72	0.68	0.14
Mut2Vec	0.49	0.57	0.63	0.57	0.18
TCGA-Embedding	0.30	0.34	0.41	0.36	0.26
SeqVec	0.84	0.89	0.91	0.88	0.05
MSA-Transformer	0.59	0.67	0.73	0.68	0.14
CPCProt	0.52	0.65	0.70	0.68	0.16
ProtBERT-BFD	0.77	0.85	0.87	0.84	0.07
TAPE-BERT-PFAM	0.78	0.85	0.86	0.85	0.07
ESM-1b	0.77	0.83	0.87	0.84	0.07
ProtALBERT	0.83	0.89	0.91	0.87	0.05
ProtXLNet	0.71	0.82	0.85	0.82	0.09
UniRep	0.75	0.82	0.86	0.82	0.08
ProtT5-XL	**0.86**	**0.90**	0.92	**0.89**	**0.04**

(b)

Methods	**Accuracy**	**F1-Weighted**	**Precision**	**Recall**	**Hamming**
InterPro2GO	0.07	0.11	0.50	0.07	0.20
UniRule2GO	0.01	0.01	0.19	0.01	0.21
Ensembl-Orthology	0.15	0.24	**0.72**	0.16	0.18
BLAST	0.46	0.56	0.68	0.50	0.15
HMMER	0.50	0.61	0.64	0.60	0.16
K-Sep	0.36	0.52	0.59	0.55	0.23
APAAC	0.22	0.34	0.42	0.41	0.28
PFAM	0.48	0.56	0.67	0.52	0.16
AAC	0.16	0.19	0.32	0.20	0.25
ProtVec	0.24	0.36	0.41	0.42	0.27
Gene2Vec	0.31	0.44	0.49	0.42	0.23
Learned-Vec	0.31	0.39	0.48	0.39	0.22
Mut2Vec	0.33	0.43	0.49	0.42	0.22
TCGA-Embedding	0.24	0.32	0.42	0.32	0.24
SeqVec	0.50	0.60	0.65	0.59	0.16
MSA-Transformer	0.32	0.47	0.57	0.52	0.23
CPCProt	0.27	0.40	0.51	0.41	0.25
ProtBERT-BFD	0.49	0.61	0.70	0.59	0.16
TAPE-BERT-PFAM	0.40	0.54	0.61	0.55	0.20
ESM-1b	0.42	0.53	0.64	0.55	0.19
ProtALBERT	0.51	0.63	0.71	0.60	0.15
ProtXLNet	0.37	0.50	0.61	0.51	0.20
UniRep	0.38	0.48	0.60	0.49	0.21
ProtT5-XL	**0.57**	**0.66**	0.71	**0.63**	**0.13**

(c)

Methods	Accuracy	F1-Weighted	Precision	Recall	Hamming
InterPro2GO	0.22	0.27	0.72	0.22	0.18
UniRule2GO	0.03	0.04	0.17	0.03	0.22
Ensembl-Orthology	0.16	0.26	0.72	0.17	0.19
BLAST	0.49	0.57	0.70	0.53	0.14
HMMER	0.49	0.60	0.65	0.59	0.16
K-Sep	0.34	0.50	0.58	0.54	0.21
APAAC	0.24	0.40	0.45	0.46	0.26
PFAM	0.50	0.58	0.71	0.54	0.13
AAC	0.24	0.23	0.28	0.28	0.23
ProtVec	0.24	0.38	0.45	0.42	0.26
Gene2Vec	0.37	0.50	0.55	0.50	0.19
Learned-Vec	0.32	0.41	0.47	0.42	0.21
Mut2Vec	0.36	0.46	0.53	0.46	0.19
TCGA-Embedding	0.34	0.41	0.50	0.44	0.21
SeqVec	0.50	0.61	0.68	0.59	0.15
MSA-Transformer	0.34	0.50	0.59	0.52	0.21
CPCProt	0.32	0.44	0.52	0.46	0.21
ProtBERT-BFD	0.51	0.62	0.69	0.62	0.15
TAPE-BERT-PFAM	0.45	0.58	0.66	0.58	0.17
ESM-1b	0.47	0.61	0.70	0.63	0.16
ProtALBERT	0.51	0.64	0.73	0.61	0.14
ProtXLNet	0.45	0.59	0.67	0.60	0.17
UniRep	0.38	0.53	0.59	0.57	0.19
ProtT5-XL	**0.59**	**0.68**	**0.76**	**0.66**	**0.12**

Table S5. Average performance results (weighted F1-score) in terms of GO groups such as “low”, “middle”, “high”, and “specific”, “normal”, “shallow”, in ontology-based protein function prediction benchmark for GO: **(a)** molecular function, **(b)** biological process, **(c)** cellular component. “*” sign denotes the method performs better than BLAST (considering all categories) and the difference is statistically significant (FDR<0.05).

(a)

	GO category						
Methods	**Low**	**Middle**	**High**	**Specific**	**Normal**	**Shallow**	**Methods Mean**
InterPro2GO	0.39	0.47	0.18	0.11	0.46	0.45	0.35
UniRule2GO	0.01	0.01	0.01	0.00	0.02	0.01	0.01
Ensembl-Orthology	0.15	0.25	0.20	0.14	0.21	0.23	0.20
BLAST	0.86	0.89	0.88	0.90	0.91	0.82	0.88
HMMER	**0.89**	0.89	0.88	**0.92**	0.91	0.85	0.89
K-Sep	0.82	0.81	0.79	0.84	0.85	0.75	0.81
APAAC	0.65	0.49	0.62	0.51	0.70	0.51	0.58
PFAM	0.85	0.86	0.88	0.84	0.89	0.86	0.86
AAC	0.37	0.36	0.55	0.26	0.59	0.33	0.41
ProtVec	0.67	0.55	0.71	0.57	0.72	0.59	0.64
Gene2Vec	0.52	0.60	0.45	0.48	0.65	0.44	0.52
Learned-Vec	0.66	0.66	0.71	0.63	0.75	0.63	0.68
Mut2Vec	0.50	0.60	0.64	0.54	0.65	0.52	0.58
TCGA-Embedding	0.36	0.35	0.30	0.35	0.48	0.20	0.34
SeqVec	0.86	**0.91**	0.90	0.91	0.91	0.85	0.89
MSA-Transformer	0.55	0.71	0.78	0.72	0.71	0.59	0.68
CPCProt	0.64	0.61	0.74	0.61	0.75	0.59	0.66
ProtBERT-BFD	0.82	0.86	0.88	0.85	0.85	0.84	0.85
TAPE-BERT-PFAM	0.83	0.86	0.85	0.86	0.87	0.82	0.85
ESM-1b	0.80	0.85	0.85	0.85	0.89	0.76	0.83
ProtALBERT	0.86	0.90	0.90	0.89	0.90	**0.87**	0.89
ProtXLNet	0.79	0.82	0.84	0.84	0.84	0.78	0.82
UniRep	0.85	0.80	0.83	0.86	0.87	0.75	0.83
ProtT5-XL	**0.89**	**0.91**	**0.92**	**0.92**	**0.93**	0.86	**0.90**
Category Mean:	0.65	0.67	0.68	0.64	0.72	0.62	0.66

(b)

Methods	GO category						Method Mean
	Low	Middle	High	Specific	Normal	Shallow	
InterPro2GO	0.18	0.10	0.06	0.08	0.15	0.10	0.11
UniRule2GO	0.03	0.00	0.00	0.02	0.01	0.00	0.01
Ensembl-Orthology	0.23	0.36	0.13	0.25	0.27	0.20	0.24
BLAST	0.36	0.68	0.63	0.63	0.48	0.56	0.56
HMMER*	**0.52**	0.69	0.63	0.67	0.56	0.60	0.61
K-Sep	0.47	0.60	0.48	0.61	0.46	0.48	0.52
APAAC	0.29	0.41	0.34	0.46	0.26	0.32	0.34
PFAM	0.37	0.67	0.63	0.62	0.52	0.53	0.56
AAC	0.15	0.27	0.14	0.25	0.13	0.18	0.19
ProtVec	0.32	0.42	0.34	0.51	0.30	0.28	0.36
Gene2Vec	0.38	0.52	0.41	0.47	0.42	0.42	0.44
Learned-Vec	0.40	0.46	0.32	0.55	0.28	0.35	0.39
Mut2Vec	0.29	0.49	0.50	0.52	0.36	0.40	0.43
TCGA-Embedding	0.37	0.37	0.22	0.37	0.31	0.28	0.32
SeqVec	0.45	0.73	0.62	0.68	0.55	0.57	0.60
MSA-Transformer	0.38	0.53	0.50	0.57	0.41	0.43	0.47
CPCProt	0.32	0.48	0.39	0.48	0.33	0.37	0.40
ProtBERT-BFD	**0.52**	0.69	0.62	0.69	0.58	0.57	0.61
TAPE-BERT-PFAM	0.48	0.63	0.50	0.65	0.51	0.44	0.54
ESM-1b	0.35	0.65	0.59	0.58	0.52	0.49	0.53
ProtALBERT*	0.47	0.74	**0.67**	0.65	**0.64**	0.59	0.63
ProtXLNet	0.33	0.60	0.56	0.58	0.48	0.44	0.50
UniRep	0.42	0.59	0.43	0.59	0.44	0.41	0.48
ProtT5-XL*	**0.52**	**0.78**	**0.67**	**0.72**	0.60	**0.65**	**0.66**
Category Mean:	0.36	0.52	0.43	0.51	0.40	0.40	0.44

(c)

Methods	GO category						Method Mean
	Low	Middle	High	Specific	Normal	Shallow	
InterPro2GO	0.34	0.14	0.37	0.17	0.16	0.45	0.27
UniRule2GO	0.00	0.05	0.08	0.04	0.05	0.03	0.04
Ensembl-Orthology	0.23	0.26	0.32	0.20	0.27	0.30	0.26
BLAST	0.46	0.59	0.71	0.44	0.64	0.59	0.57
HMMER	0.52	0.61	0.69	0.47	0.65	0.63	0.59
K-Sep	0.41	0.56	0.56	0.48	0.50	0.52	0.50
APAAC	0.35	0.41	0.44	0.35	0.39	0.43	0.40
PFAM	0.47	0.61	0.70	0.45	0.64	0.61	0.58
AAC	0.13	0.21	0.43	0.11	0.27	0.27	0.24
ProtVec	0.26	0.42	0.50	0.26	0.41	0.43	0.38
Gene2Vec	0.43	0.56	0.52	0.46	0.52	0.50	0.50
Learned-Vec	0.30	0.46	0.50	0.28	0.44	0.46	0.41
Mut2Vec	0.37	0.50	0.54	0.47	0.44	0.48	0.47
TCGA-Embedding	0.35	0.43	0.47	0.33	0.49	0.39	0.41
SeqVec	0.49	0.66	0.70	0.49	0.62	0.67	0.61
MSA-Transformer	0.36	0.55	0.64	0.36	0.57	0.54	0.50
CPCProt	0.32	0.47	0.55	0.35	0.47	0.46	0.44
ProtBERT-BFD	0.53	0.67	0.69	0.49	0.67	0.67	0.62
TAPE-BERT-PFAM	0.50	0.63	0.62	0.51	0.61	0.60	0.58
ESM-1b	0.51	0.64	0.71	0.54	0.60	0.67	0.61
ProtALBERT*	0.54	0.67	0.73	0.51	0.66	0.70	0.63
ProtXLNet	0.50	0.63	0.66	0.47	0.61	0.64	0.59
UniRep	0.45	0.56	0.61	0.41	0.58	0.57	0.53
ProtT5-XL*	**0.58**	**0.74**	**0.74**	**0.57**	**0.69**	**0.75**	**0.68**
Category Mean:	0.39	0.50	0.56	0.38	0.50	0.51	0.48

Table S6. Statistical significance corresponding to the performance differences between representation methods in the ontology-based protein function prediction benchmark. FDR values are calculated using the Wilcoxon Rank Sum test with Benjamini-Hochberg correction for GO: **(a)** molecular function, **(b)** biological process, **(c)** cellular component. Color codes represent FDR values (pink: FDR > 0.05, yellow: 0.05 > FDR > 0.0005, green: FDR =< 0.0005).

(a)

	AAC	ProtALBERT	APAAC	ProtBERT-BFD	TAPE-BERT-PFAM	BLAST	CPCProt	ESM-1b	Gene2Vec	HMMER	K-Sep	Learned-Vec	MSA-Transformer	Mut2Vec	PFAM	ProtVec	SeqVec	ProtT5-XL	TCGA-Embedding	UniRep	ProtXLNet
AAC	1.0E+0	1.9E-7	6.2E-6	2.4E-7	1.9E-7	1.9E-7	5.1E-7	1.9E-7	3.0E-3	1.9E-7	1.9E-7	1.1E-6	6.3E-5	9.5E-5	1.9E-7	5.2E-7	1.9E-7	1.9E-7	7.3E-2	1.9E-7	1.9E-7
ProtALBERT	1.9E-7	1.0E+0	1.9E-7	1.2E-2	4.8E-3	5.1E-1	4.6E-7	2.5E-4	1.9E-7	9.7E-1	2.0E-4	2.8E-7	4.4E-7	1.9E-7	8.0E-2	3.1E-7	1.0E+0	4.3E-1	1.9E-7	1.9E-3	8.2E-6
APAAC	6.2E-6	1.9E-7	1.0E+0	8.0E-7	2.1E-7	2.6E-7	4.4E-3	1.9E-7	1.0E-1	1.9E-7	5.9E-7	1.8E-3	1.5E-2	7.3E-1	1.9E-7	1.5E-2	2.0E-7	1.9E-7	3.1E-6	1.9E-7	4.8E-7
ProtBERT-BFD	2.4E-7	1.2E-2	8.0E-7	1.0E+0	1.7E-1	5.1E-2	4.2E-6	6.6E-2	2.5E-7	7.1E-3	8.2E-3	2.3E-6	3.5E-6	8.2E-7	5.4E-1	3.0E-6	1.4E-3	1.6E-3	1.9E-7	3.4E-2	1.6E-3
TAPE-BERT-PFAM	1.9E-7	4.8E-3	2.1E-7	1.7E-1	1.0E+0	3.1E-2	9.1E-7	5.5E-1	1.9E-7	2.0E-3	2.0E-2	1.9E-7	8.7E-7	1.9E-7	3.3E-1	1.9E-7	2.1E-4	1.1E-4	1.9E-7	2.6E-1	2.6E-2
BLAST	1.9E-7	5.1E-1	2.6E-7	5.1E-2	3.1E-2	1.0E+0	7.8E-7	2.0E-3	1.9E-7	2.7E-1	2.1E-4	3.8E-7	2.5E-7	1.9E-7	1.9E-1	2.9E-7	3.1E-1	6.1E-2	1.9E-7	7.3E-3	9.0E-5
CPCProt	5.1E-7	4.6E-7	4.4E-3	4.2E-6	9.1E-7	7.8E-7	1.0E+0	3.3E-7	2.1E-4	2.4E-7	7.8E-7	3.3E-1	8.7E-2	9.5E-4	2.7E-7	2.0E-1	3.2E-7	3.3E-7	2.1E-7	3.8E-7	1.1E-6
ESM-1b	1.9E-7	2.5E-4	1.9E-7	6.6E-2	5.5E-1	2.0E-3	3.3E-7	1.0E+0	1.9E-7	4.9E-5	1.1E-1	1.0E-6	2.4E-6	1.9E-7	5.3E-2	2.4E-7	1.9E-4	6.4E-6	1.9E-7	4.4E-1	6.1E-2
Gene2Vec	3.0E-3	1.9E-7	1.0E-1	2.5E-7	1.9E-7	1.9E-7	2.1E-4	1.9E-7	1.0E+0	1.9E-7	1.9E-7	6.5E-5	3.5E-3	1.3E-1	1.9E-7	4.1E-3	1.9E-7	1.9E-7	3.8E-6	1.9E-7	1.9E-7
HMMER	1.9E-7	9.7E-1	1.9E-7	7.1E-3	2.0E-3	2.7E-1	2.4E-7	4.9E-5	1.9E-7	1.0E+0	4.2E-6	1.9E-7	2.0E-7	1.9E-7	5.5E-3	1.9E-7	6.2E-1	4.2E-1	1.9E-7	2.5E-5	6.2E-6
K-Sep	1.9E-7	2.0E-4	5.9E-7	8.2E-3	2.0E-2	2.1E-4	7.8E-7	1.1E-1	1.9E-7	4.2E-6	1.0E+0	8.9E-7	1.2E-4	1.9E-7	2.6E-3	7.7E-7	8.6E-5	3.2E-6	1.9E-7	4.8E-1	5.6E-1
Learned-Vec	1.1E-6	2.8E-7	1.8E-3	2.3E-6	1.9E-7	3.8E-7	3.3E-1	1.0E-6	6.5E-5	1.9E-7	8.9E-7	1.0E+0	5.0E-1	1.0E-4	2.4E-7	1.6E-1	2.4E-7	2.2E-7	7.0E-7	1.9E-7	1.4E-6
MSA-Transformer	6.3E-5	4.4E-7	1.5E-2	3.5E-6	8.7E-7	2.5E-7	8.7E-2	2.4E-6	3.5E-3	2.0E-7	1.2E-4	5.0E-1	1.0E+0	4.5E-3	6.5E-7	1.0E-1	2.0E-7	2.0E-7	1.1E-5	1.7E-6	2.7E-6
Mut2Vec	9.5E-5	1.9E-7	7.3E-1	8.2E-7	1.9E-7	1.9E-7	9.5E-4	1.9E-7	1.3E-1	1.9E-7	1.9E-7	1.0E-4	4.5E-3	1.0E+0	1.9E-7	6.0E-2	1.9E-7	1.9E-7	6.1E-6	1.9E-7	1.9E-7
PFAM	1.9E-7	8.0E-2	1.9E-7	5.4E-1	3.3E-1	1.9E-1	2.7E-7	5.3E-2	1.9E-7	5.5E-3	2.6E-3	2.4E-7	6.5E-7	1.9E-7	1.0E+0	2.4E-7	7.3E-2	6.6E-3	1.9E-7	1.9E-2	9.1E-4
ProtVec	5.2E-7	3.1E-7	1.5E-2	3.0E-6	1.9E-7	2.9E-7	2.0E-1	2.4E-7	4.1E-3	1.9E-7	7.7E-7	1.6E-1	1.0E-1	6.0E-2	2.4E-7	1.0E+0	2.7E-7	1.9E-7	1.3E-6	1.9E-7	8.7E-7
SeqVec	1.9E-7	1.0E+0	2.0E-7	1.4E-3	2.1E-4	3.1E-1	3.2E-7	1.9E-4	1.9E-7	6.2E-1	8.6E-5	2.4E-7	2.0E-7	1.9E-7	7.3E-2	2.7E-7	1.0E+0	1.3E-1	1.9E-7	8.3E-4	1.1E-5
ProtT5-XL	1.9E-7	4.3E-1	1.9E-7	1.6E-3	1.1E-4	6.1E-2	3.3E-7	6.4E-6	1.9E-7	4.2E-1	3.2E-6	2.2E-7	2.0E-7	1.9E-7	6.6E-3	1.9E-7	1.3E-1	1.0E+0	1.9E-7	6.7E-5	2.9E-7
TCGA-Embedding	7.3E-2	1.9E-7	3.1E-6	1.9E-7	1.9E-7	1.9E-7	2.1E-7	1.9E-7	3.8E-6	1.9E-7	1.9E-7	7.0E-7	1.1E-5	6.1E-6	1.9E-7	1.3E-6	1.9E-7	1.9E-7	1.0E+0	1.9E-7	1.9E-7
UniRep	1.9E-7	1.9E-3	1.9E-7	3.4E-2	2.6E-1	7.3E-3	3.8E-7	4.4E-1	1.9E-7	2.5E-5	4.8E-1	1.9E-7	1.7E-6	1.9E-7	1.9E-2	1.9E-7	8.3E-4	6.7E-5	1.9E-7	1.0E+0	6.2E-1
ProtXLNet	1.9E-7	8.2E-6	4.8E-7	1.6E-3	2.6E-2	9.0E-5	1.1E-6	6.1E-2	1.9E-7	6.2E-6	5.6E-1	1.4E-6	2.7E-6	1.9E-7	9.1E-4	8.7E-7	1.1E-5	2.9E-7	1.9E-7	6.2E-1	1.0E+0

(b)

	AAC	ProtALBERT	APAAC	ProtBERT-BFD	TAPE-BERT-PFAM	BLAST	CPCProt	ESM-1b	Gene2Vec	HMMER	K-Sep	Learned-Vec	MSA-Transformer	Mut2Vec	PFAM	ProtVec	SeqVec	ProtT5-XL	TCGA-Embedding	UniRep	ProtXLNet
AAC	1.0E+0	1.7E-7	5.9E-6	1.7E-7	1.7E-7	2.9E-7	7.8E-7	1.7E-7	4.6E-7	1.7E-7	1.7E-7	3.3E-7	2.5E-7	5.0E-7	3.0E-7	1.1E-6	1.7E-7	1.7E-7	1.4E-4	1.7E-7	4.4E-7
ProtALBERT	1.7E-7	1.0E+0	9.9E-7	5.9E-2	2.1E-3	2.5E-3	8.7E-7	4.4E-5	2.5E-5	7.1E-2	3.8E-4	3.7E-6	1.6E-5	4.4E-6	1.4E-3	2.1E-6	6.0E-2	7.9E-2	1.5E-6	1.2E-4	5.0E-6
APAAC	5.9E-6	9.9E-7	1.0E+0	8.7E-7	2.1E-6	1.3E-5	7.1E-2	4.4E-6	2.1E-3	6.5E-7	1.4E-6	2.3E-1	4.0E-4	9.6E-4	2.0E-5	3.4E-1	4.0E-7	5.0E-7	1.7E-1	3.9E-5	1.3E-4
ProtBERT-BFD	1.7E-7	5.9E-2	8.7E-7	1.0E+0	1.9E-3	2.4E-1	4.6E-7	1.3E-3	8.7E-5	4.6E-1	1.2E-4	2.4E-6	1.6E-6	1.4E-6	2.4E-1	8.7E-7	6.8E-1	2.7E-3	1.5E-6	1.6E-5	9.1E-6
TAPE-BERT-PFAM	1.7E-7	2.1E-3	2.1E-6	1.9E-3	1.0E+0	2.1E-1	9.7E-6	3.4E-1	2.1E-3	8.0E-3	3.0E-1	6.1E-5	1.3E-2	2.0E-3	2.2E-1	2.5E-6	1.6E-2	1.4E-3	6.5E-5	4.2E-2	5.3E-1
BLAST	2.9E-7	2.5E-3	1.3E-5	2.4E-1	2.1E-1	1.0E+0	5.6E-5	2.7E-1	4.7E-4	5.2E-3	8.6E-2	7.8E-5	2.7E-3	1.5E-5	9.8E-1	3.3E-5	8.6E-2	2.5E-5	9.1E-6	6.4E-3	1.7E-2
CPCProt	7.8E-7	8.7E-7	7.1E-2	4.6E-7	9.7E-6	5.6E-5	1.0E+0	2.5E-5	8.4E-2	3.2E-6	2.5E-5	3.9E-1	6.2E-4	1.0E-1	1.3E-4	5.0E-2	5.1E-6	3.3E-7	1.1E-2	2.5E-3	1.3E-3
ESM-1b	1.7E-7	4.4E-5	4.4E-6	1.3E-3	3.4E-1	2.7E-1	2.5E-5	1.0E+0	4.7E-3	2.9E-3	2.4E-1	4.2E-4	9.2E-3	2.5E-5	5.3E-1	7.7E-6	1.1E-2	2.2E-6	9.0E-5	3.2E-2	3.8E-2
Gene2Vec	4.6E-7	2.5E-5	2.1E-3	8.7E-5	2.1E-3	4.7E-4	8.4E-2	4.7E-3	1.0E+0	6.1E-6	1.9E-3	7.5E-2	2.0E-1	9.4E-1	6.2E-4	1.1E-2	1.2E-4	2.0E-5	7.3E-5	2.0E-1	4.0E-2
HMMER	1.7E-7	7.1E-2	6.5E-7	4.6E-1	8.0E-3	5.2E-3	3.2E-6	2.9E-3	6.1E-6	1.0E+0	4.2E-4	2.1E-6	3.4E-6	5.0E-7	2.0E-2	6.6E-7	4.9E-1	5.3E-3	4.6E-7	3.8E-5	7.4E-5
K-Sep	1.7E-7	3.8E-4	1.4E-6	1.2E-4	3.0E-1	8.6E-2	2.5E-5	2.4E-1	1.9E-3	4.2E-4	1.0E+0	5.6E-5	6.7E-2	5.0E-3	1.4E-1	2.1E-6	1.3E-3	4.4E-5	6.3E-6	1.7E-1	7.9E-1
Learned-Vec	3.3E-7	3.7E-6	2.3E-1	2.4E-6	6.1E-5	7.8E-5	3.9E-1	4.2E-4	7.5E-2	2.1E-6	5.6E-5	1.0E+0	5.6E-3	3.2E-2	8.7E-5	6.0E-1	4.4E-6	9.0E-7	1.7E-3	2.3E-3	5.2E-3
MSA-Transformer	2.5E-7	1.6E-5	4.0E-4	1.6E-6	1.3E-2	2.7E-3	6.2E-4	9.2E-3	2.0E-1	3.4E-6	6.7E-2	5.6E-3	1.0E+0	6.5E-2	2.5E-3	1.0E-4	1.9E-5	7.8E-7	1.1E-4	7.5E-1	1.6E-1
Mut2Vec	5.0E-7	4.4E-6	9.6E-4	1.4E-6	2.0E-3	1.5E-5	1.0E-1	2.5E-5	9.4E-1	5.0E-7	5.0E-3	3.2E-2	6.5E-2	1.0E+0	2.8E-5	2.1E-3	7.4E-7	5.1E-7	2.4E-3	1.2E-1	5.7E-3
PFAM	3.0E-7	1.4E-3	2.0E-5	2.4E-1	2.2E-1	9.8E-1	1.3E-4	5.3E-1	6.2E-4	2.0E-2	1.4E-1	8.7E-5	2.5E-3	2.8E-5	1.0E+0	1.9E-5	1.6E-1	3.0E-5	1.3E-5	1.1E-2	2.2E-2
ProtVec	1.1E-6	2.1E-6	3.4E-1	8.7E-7	2.5E-6	3.3E-5	5.0E-2	7.7E-6	1.1E-2	6.6E-7	2.1E-6	6.0E-1	1.0E-4	2.1E-3	1.9E-5	1.0E+0	7.8E-7	5.0E-7	5.0E-2	6.0E-5	2.0E-4
SeqVec	1.7E-7	6.0E-2	4.0E-7	6.8E-1	1.6E-2	8.6E-2	5.1E-6	1.1E-2	1.2E-4	4.9E-1	1.3E-3	4.4E-6	1.9E-5	7.4E-7	1.6E-1	7.8E-7	1.0E+0	1.0E-2	2.1E-6	1.1E-5	3.8E-4
ProtT5-XL	1.7E-7	7.9E-2	5.0E-7	2.7E-3	1.4E-3	2.5E-5	3.3E-7	2.2E-6	2.0E-5	5.3E-3	4.4E-5	9.0E-7	7.8E-7	5.1E-7	3.0E-5	5.0E-7	1.0E-2	1.0E+0	7.8E-7	6.3E-6	1.3E-6
TCGA-Embedding	1.4E-4	1.5E-6	1.7E-1	1.5E-6	6.5E-5	9.1E-6	1.1E-2	9.0E-5	7.3E-5	4.6E-7	6.3E-6	1.7E-3	1.1E-4	2.4E-3	1.3E-5	5.0E-2	2.1E-6	7.8E-7	1.0E+0	1.7E-5	2.5E-4
UniRep	1.7E-7	1.2E-4	3.9E-5	1.6E-5	4.2E-2	6.4E-3	2.5E-3	3.2E-2	2.0E-1	3.8E-5	1.7E-1	2.3E-3	7.5E-1	1.2E-1	1.1E-2	6.0E-5	1.1E-5	6.3E-6	1.7E-5	1.0E+0	4.0E-1
ProtXLNet	4.4E-7	5.0E-6	1.3E-4	9.1E-6	5.3E-1	1.7E-2	1.3E-3	3.8E-2	4.0E-2	7.4E-5	7.9E-1	5.2E-3	1.6E-1	5.7E-3	2.2E-2	2.0E-4	3.8E-4	1.3E-6	2.5E-4	4.0E-1	1.0E+0

(c)

	AAC	ProtALBERT	APAAC	ProtBERT-BFD	TAPE-BERT-PFAM	BLAST	CPCProt	ESM-1b	Gene2Vec	HMMER	K-Sep	Learned-Vec	MSA-Transformer	Mut2Vec	PFAM	ProtVec	SeqVec	ProtT5-XL	TCGA-Embedding	UniRep	ProtXLNet
AAC	1.0E+0	4.4E-7	7.9E-6	4.4E-7	4.7E-7	4.4E-7	2.3E-6	4.4E-7	4.4E-7	4.4E-7	6.1E-7	5.8E-7	3.5E-6	5.2E-7	4.4E-7	2.7E-5	4.4E-7	4.4E-7	3.6E-6	4.4E-7	4.4E-7
ProtALBERT	4.4E-7	1.0E+0	1.5E-6	4.8E-1	8.7E-3	2.3E-2	8.2E-7	2.0E-1	2.9E-5	4.9E-2	4.2E-5	1.3E-6	2.5E-4	2.3E-5	3.1E-2	4.7E-7	1.4E-1	1.3E-3	1.8E-6	6.8E-5	2.5E-2
APAAC	7.9E-6	1.5E-6	1.0E+0	2.1E-6	8.6E-6	4.9E-6	2.9E-2	8.2E-7	5.4E-4	7.7E-7	6.2E-4	8.5E-1	5.2E-3	6.9E-3	3.9E-6	9.8E-1	7.6E-7	4.5E-7	7.3E-1	1.6E-5	2.5E-6
ProtBERT-BFD	4.4E-7	4.8E-1	2.1E-6	1.0E+0	4.6E-3	1.0E-1	1.1E-6	9.3E-1	9.7E-5	1.4E-1	2.2E-4	1.8E-6	2.3E-5	1.4E-4	9.2E-2	7.3E-7	6.6E-1	9.1E-4	2.5E-6	1.7E-4	6.5E-2
TAPE-BERT-PFAM	4.7E-7	8.7E-3	8.6E-6	4.6E-3	1.0E+0	9.1E-1	1.5E-6	6.0E-2	2.2E-3	4.8E-1	7.1E-4	2.0E-6	3.8E-3	1.3E-4	9.7E-1	8.6E-7	1.7E-1	3.2E-5	9.4E-7	3.4E-2	6.7E-1
BLAST	4.4E-7	2.3E-2	4.9E-6	1.0E-1	9.1E-1	1.0E+0	1.1E-5	1.7E-1	2.4E-3	6.7E-1	1.1E-2	1.6E-6	1.2E-2	5.0E-4	3.9E-1	4.4E-7	9.2E-2	1.6E-4	1.4E-5	3.4E-2	6.4E-1
CPCProt	2.3E-6	8.2E-7	2.9E-2	1.1E-6	1.5E-6	1.1E-5	1.0E+0	1.6E-6	1.4E-2	6.8E-7	5.0E-3	1.4E-1	7.6E-3	4.4E-1	3.6E-6	2.8E-2	1.3E-6	4.4E-7	1.6E-1	1.3E-4	7.3E-7
ESM-1b	4.4E-7	2.0E-1	8.2E-7	9.3E-1	6.0E-2	1.7E-1	1.6E-6	1.0E+0	7.8E-4	1.9E-1	5.7E-4	2.1E-6	2.2E-3	4.6E-6	2.2E-1	5.8E-7	6.0E-1	4.6E-4	4.8E-6	7.8E-3	1.8E-1
Gene2Vec	4.4E-7	2.9E-5	5.4E-4	9.7E-5	2.2E-3	2.4E-3	1.4E-2	7.8E-4	1.0E+0	1.6E-4	5.5E-1	1.7E-4	8.5E-1	8.0E-2	8.5E-4	1.9E-5	7.4E-5	6.0E-6	4.1E-4	1.1E-1	1.1E-3
HMMER	4.4E-7	4.9E-2	7.7E-7	1.4E-1	4.8E-1	6.7E-1	6.8E-7	1.9E-1	1.6E-4	1.0E+0	6.5E-4	4.4E-7	2.4E-3	3.6E-5	1.0E+0	4.4E-7	2.4E-1	2.9E-4	1.6E-6	2.7E-3	9.2E-1
K-Sep	6.1E-7	4.2E-5	6.2E-4	2.2E-4	7.1E-4	1.1E-2	5.0E-3	5.7E-4	5.5E-1	6.5E-4	1.0E+0	4.3E-4	1.0E+0	1.4E-1	2.1E-3	5.3E-5	1.3E-5	2.5E-6	2.2E-3	2.0E-1	7.1E-4
Learned-Vec	5.8E-7	1.3E-6	8.5E-1	1.8E-6	2.0E-6	1.6E-6	1.4E-1	2.1E-6	1.7E-4	4.4E-7	4.3E-4	1.0E+0	8.9E-4	1.9E-2	9.0E-7	2.1E-1	4.9E-7	4.4E-7	1.0E+0	1.9E-5	2.3E-6
MSA-Transformer	3.5E-6	2.5E-4	5.2E-3	2.3E-5	3.8E-3	1.2E-2	7.6E-3	2.2E-3	8.5E-1	2.4E-3	1.0E+0	8.9E-4	1.0E+0	1.3E-1	5.2E-3	5.0E-5	6.6E-5	1.8E-6	3.5E-3	6.0E-1	6.7E-3
Mut2Vec	5.2E-7	2.3E-5	6.9E-3	1.4E-4	1.3E-4	5.0E-4	4.4E-1	4.6E-6	8.0E-2	3.6E-5	1.4E-1	1.9E-2	1.3E-1	1.0E+0	1.5E-4	2.3E-3	5.3E-5	1.6E-6	1.4E-2	5.6E-3	2.0E-4
PFAM	4.4E-7	3.1E-2	3.9E-6	9.2E-2	9.7E-1	3.9E-1	3.6E-6	2.2E-1	8.5E-4	1.0E+0	2.1E-3	9.0E-7	5.2E-3	1.5E-4	1.0E+0	4.4E-7	4.7E-2	1.3E-4	6.0E-6	3.9E-3	1.0E+0
ProtVec	2.7E-5	4.7E-7	9.8E-1	7.3E-7	8.6E-7	4.4E-7	2.8E-2	5.8E-7	1.9E-5	4.4E-7	5.3E-5	2.1E-1	5.0E-5	2.3E-3	4.4E-7	1.0E+0	4.4E-7	4.4E-7	3.0E-1	1.6E-6	8.2E-7
SeqVec	4.4E-7	1.4E-1	7.6E-7	6.6E-1	1.7E-1	9.2E-2	1.3E-6	6.0E-1	7.4E-5	2.4E-1	1.3E-5	4.9E-7	6.6E-5	5.3E-5	4.7E-2	4.4E-7	1.0E+0	1.3E-4	1.8E-6	2.4E-4	2.1E-1
ProtT5-XL	4.4E-7	1.3E-3	4.5E-7	9.1E-4	3.2E-5	1.6E-4	4.4E-7	4.6E-4	6.0E-6	2.9E-4	2.5E-6	4.4E-7	1.8E-6	1.6E-6	1.3E-4	4.4E-7	1.3E-4	1.0E+0	7.7E-7	3.3E-6	5.7E-6
TCGA-Embedding	3.6E-6	1.8E-6	7.3E-1	2.5E-6	9.4E-7	1.4E-5	1.6E-1	4.8E-6	4.1E-4	1.6E-6	2.2E-3	1.0E+0	3.5E-3	1.4E-2	6.0E-6	3.0E-1	1.8E-6	7.7E-7	1.0E+0	2.1E-4	1.5E-5
UniRep	4.4E-7	6.8E-5	1.6E-5	1.7E-4	3.4E-2	3.4E-2	1.3E-4	7.8E-3	1.1E-1	2.7E-3	2.0E-1	1.9E-5	6.0E-1	5.6E-3	3.9E-3	1.6E-6	2.4E-4	3.3E-6	2.1E-4	1.0E+0	4.6E-3
ProtXLNet	4.4E-7	2.5E-2	2.5E-6	6.5E-2	6.7E-1	6.4E-1	7.3E-7	1.8E-1	1.1E-3	9.2E-1	7.1E-4	2.3E-6	6.7E-3	2.0E-4	1.0E+0	8.2E-7	2.1E-1	5.7E-6	1.5E-5	4.6E-3	1.0E+0

Table S7. Overall sample statistics of the dataset used in drug target protein family classification benchmark, per target protein family and representation model.

Method name	Enzymes	Membrane receptors	Transcription factors	Ion channels	Others	Total
BLAST	4,360	831	1,034	347	1,019	7,591
HMMER	4,360	831	1,034	347	1,019	7,591
K-Sep	4,339	830	1,033	346	1,015	7,563
APAAC	4,360	831	1,034	347	1,019	7,591
PFAM	4,360	831	1,034	347	1,019	7,591
AAC	4,360	831	1,034	347	1,019	7,591
ProtVec	4,360	831	1,034	347	1,019	7,591
Gene2Vec	4,192	470	990	336	1,008	6,996
Learned-Vec	4,360	831	1,034	347	1,019	7,591
Mut2Vec	4,119	718	983	334	965	7,119
TCGA-Embedding	4,265	766	1,032	344	1,010	7,417
SeqVec	4,360	831	1,034	347	1,019	7,591
MSA-Transformer	4342	829	1029	343	1005	7548
CPCProt	4,284	825	1,032	328	992	7,461
ProtBERT-BFD	4,284	825	1,032	328	992	7,461
TAPE-BERT-PFAM	4,285	825	1,032	328	992	7,462
ESM-1b	4,284	825	1,032	328	992	7,461
ProtALBERT	4,284	825	1,032	328	992	7,461
ProtXLNet	4,284	825	1,032	328	992	7,461
UniRep	4,360	831	1,034	347	1,019	7,591
ProtT5-XL	4,284	825	1,032	328	992	7,461

Table S8. Results of the statistical significance test between representation methods used in the drug target protein family classification benchmark. FDR values are calculated using the Wilcoxon Rank Sum test with Benjamini-Hochberg correction. FDR values are based on results of **(a)** "random split" dataset, **(b)** "UniClust50" dataset, **(c)** "UniClust30" dataset, **(d)** "MMSEQ-15" dataset. Color codes represent FDR values (pink: FDR > 0.05, yellow: 0.05 > FDR > 0.0005, green: FDR =< 0.0005).

(a)

	APAAC	TCGA-Embedding	Gene2Vec^	ProtVec	Mut2Vec^	AAC	Learned-Vec	CPCProt	K-Sep	MSA-Transformer	UniRep	TAPE-BERT-PFAM^	ProtXLNet	ProtBERT-BFD	HMMER	BLAST	ESM-1b	SeqVec	PFAM^	ProtT5-XL	ProtALBERT
APAAC	1.0E+0	3.7E-1	5.5E-1	6.1E-1	2.5E-2	8.2E-3	3.1E-3	3.1E-3	5.7E-3	3.1E-3	3.1E-3	5.7E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
TCGA-Embedding	3.7E-1	1.0E+0	8.3E-1	6.8E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
Gene2Vec^	5.5E-1	8.3E-1	1.0E+0	6.1E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
ProtVec	6.1E-1	6.8E-1	6.1E-1	1.0E+0	2.3E-1	2.5E-2	3.1E-3	3.1E-3	8.2E-3	3.1E-3	5.7E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
Mut2Vec^	2.5E-2	3.1E-3	3.1E-3	2.3E-1	1.0E+0	4.7E-2	3.1E-3	3.1E-3	2.5E-2	1.3E-2	8.2E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
AAC	8.2E-3	3.1E-3	3.1E-3	2.5E-2	4.7E-2	1.0E+0	5.7E-3	3.1E-3	2.5E-2	5.7E-3	8.2E-3	5.7E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
Learned-Vec	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.7E-3	1.0E+0	1.8E-2	3.2E-1	1.3E-1	1.3E-1	8.2E-3	5.7E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
CPCProt	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.8E-2	1.0E+0	6.1E-1	4.3E-1	2.7E-1	1.8E-2	5.7E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
K-Sep	5.7E-3	3.1E-3	3.1E-3	8.2E-3	2.5E-2	2.5E-2	3.2E-1	6.1E-1	1.0E+0	1.0E+0	1.0E+0	1.9E-1	3.5E-2	3.1E-3	3.1E-3	3.1E-3	1.3E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3
MSA-Transformer	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.3E-2	5.7E-3	1.3E-1	4.3E-1	1.0E+0	1.0E+0	9.8E-1	3.2E-1	5.7E-3	3.1E-3	5.7E-3	5.7E-3	1.3E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3
UniRep	3.1E-3	3.1E-3	3.1E-3	5.7E-3	8.2E-3	8.2E-3	1.3E-1	2.7E-1	1.0E+0	9.8E-1	1.0E+0	1.3E-1	1.0E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
TAPE-BERT-PFAM^	5.7E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.7E-3	8.2E-3	1.8E-2	1.9E-1	3.2E-1	1.3E-1	1.0E+0	2.7E-1	3.5E-2	1.3E-2	1.8E-2	3.5E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3
ProtXLNet	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.7E-3	5.7E-3	3.5E-2	5.7E-3	1.0E-1	2.7E-1	1.0E+0	6.8E-1	4.9E-1	4.3E-1	1.3E-1	1.8E-2	8.2E-3	3.1E-3	3.1E-3
ProtBERT-BFD	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.5E-2	6.8E-1	1.0E+0	3.7E-1	7.5E-1	2.3E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3
HMMER	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.7E-3	3.1E-3	1.3E-2	4.9E-1	3.7E-1	1.0E+0	1.0E+0	1.6E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3
BLAST	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.7E-3	3.1E-3	1.8E-2	4.3E-1	7.5E-1	1.0E+0	1.0E+0	1.9E-1	2.5E-2	8.2E-3	3.1E-3	3.1E-3
ESM-1b	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.3E-2	1.3E-2	3.1E-3	3.5E-2	1.3E-1	2.3E-1	1.6E-1	1.9E-1	1.0E+0	9.1E-1	4.9E-1	1.0E-1	4.7E-2
SeqVec	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.8E-2	3.1E-3	3.1E-3	2.5E-2	9.1E-1	1.0E+0	3.1E-3	3.1E-3	3.1E-3
PFAM^	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	8.2E-3	3.1E-3	3.1E-3	8.2E-3	4.9E-1	3.1E-3	1.0E+0	5.7E-3	3.1E-3
ProtT5-XL	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.0E-1	3.1E-3	5.7E-3	1.0E+0	5.4E-1
ProtALBERT	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	4.7E-2	3.1E-3	3.1E-3	5.4E-1	1.0E+0

(b)

	APAAC	ProtVec	Gene2Vec^	TCGA-Embedding	AAC	Mut2Vec^	Learned-Vec	CPCProt	MSA-Transformer	K-Sep	UniRep	TAPE-BERT-PFAM^	ProtXLNet	BLAST	ProtBERT-BFD	ESM-1b	HMMER	SeqVec	PFAM^	ProtALBERT	ProtT5-XL
APAAC	1.0E+0	1.3E-1	6.4E-2	1.1E-1	2.6E-2	3.7E-2	3.1E-3	3.1E-3	5.9E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
ProtVec	1.3E-1	1.0E+0	9.7E-1	7.5E-1	2.7E-1	1.6E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
Gene2Vec^	6.4E-2	9.7E-1	1.0E+0	8.2E-1	8.2E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
TCGA-Embedding	1.1E-1	7.5E-1	8.2E-1	1.0E+0	2.0E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
AAC	2.6E-2	2.7E-1	8.2E-2	2.0E-1	1.0E+0	1.0E+0	2.6E-2	3.1E-3	3.1E-3	1.4E-2	3.1E-3	3.1E-3	5.9E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
Mut2Vec^	3.7E-2	1.6E-1	3.1E-3	3.1E-3	1.0E+0	1.0E+0	3.1E-3	3.1E-3	5.9E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
Learned-Vec	3.1E-3	3.1E-3	3.1E-3	3.1E-3	2.6E-2	3.1E-3	1.0E+0	1.4E-2	2.6E-2	1.9E-2	3.1E-3	3.1E-3	5.9E-3	3.1E-3	3.1E-3	8.6E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
CPCProt	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.4E-2	1.0E+0	2.0E-1	2.3E-1	2.6E-2	5.9E-3	2.6E-2	3.1E-3	3.1E-3	1.4E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
MSA-Transformer	5.9E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.9E-3	2.6E-2	2.0E-1	1.0E+0	7.5E-1	6.1E-1	3.7E-1	2.7E-1	1.4E-2	8.6E-3	8.2E-2	5.9E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
K-Sep	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.4E-2	3.1E-3	1.9E-2	2.3E-1	7.5E-1	1.0E+0	7.5E-1	4.9E-2	3.2E-1	3.1E-3	3.1E-3	2.3E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
UniRep	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	2.6E-2	6.1E-1	7.5E-1	1.0E+0	1.6E-1	2.7E-1	3.1E-3	3.1E-3	4.9E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3
TAPE-BERT-PFAM^	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.9E-3	3.7E-1	4.9E-2	1.6E-1	1.0E+0	8.2E-1	3.2E-1	4.9E-2	4.3E-1	2.6E-2	3.1E-3	3.1E-3	3.1E-3	3.1E-3
ProtXLNet	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.9E-3	3.1E-3	5.9E-3	2.6E-2	2.7E-1	3.2E-1	2.7E-1	8.2E-1	1.0E+0	6.1E-1	6.1E-1	2.7E-1	6.1E-1	5.5E-1	8.2E-2	1.1E-2	3.1E-3
BLAST	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.4E-2	3.1E-3	3.1E-3	3.2E-1	6.1E-1	1.0E+0	2.3E-1	4.9E-1	1.1E-1	8.6E-3	3.1E-3	3.1E-3	3.1E-3
ProtBERT-BFD	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	8.6E-3	3.1E-3	3.1E-3	4.9E-2	6.1E-1	2.3E-1	1.0E+0	4.9E-1	9.7E-1	3.1E-3	3.1E-3	3.1E-3	3.1E-3
ESM-1b	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	8.6E-3	1.4E-2	8.2E-2	2.3E-1	4.9E-2	4.3E-1	2.7E-1	4.9E-1	4.9E-1	1.0E+0	4.9E-1	6.9E-1	9.7E-1	8.2E-2	6.4E-2
HMMER	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.9E-3	3.1E-3	3.1E-3	2.6E-2	6.1E-1	1.1E-1	9.7E-1	4.9E-1	1.0E+0	3.1E-3	3.1E-3	3.1E-3	3.1E-3
SeqVec	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	5.5E-1	8.6E-3	3.1E-3	6.9E-1	3.1E-3	1.0E+0	2.1E-2	3.1E-3	3.1E-3
PFAM^	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	8.2E-2	3.1E-3	3.1E-3	9.7E-1	3.1E-3	2.1E-2	1.0E+0	8.6E-3	3.1E-3
ProtALBERT	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	1.1E-2	3.1E-3	3.1E-3	8.2E-2	3.1E-3	3.1E-3	8.6E-3	1.0E+0	6.6E-2
ProtT5-XL	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	3.1E-3	6.4E-2	3.1E-3	3.1E-3	3.1E-3	6.6E-2	1.0E+0

(c)

	TCGA-Embedding	Gene2Vec^	APAAC	ProtVec	AAC	Mut2Vec^	Learned-Vec	CPCProt	K-Sep	MSA-Transformer	UniRep	TAPE-BERT-PFAM^	BLAST	HMMER	ProtBERT-BFD	ProtXLNet	SeqVec	PFAM^	ESM-1b	ProtALBERT	ProtT5-XL
TCGA-Embedding	1.0E+0	2.7E-1	3.2E-1	1.9E-1	7.8E-2	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
Gene2Vec^	2.7E-1	1.0E+0	3.6E-1	3.6E-1	1.0E-1	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
APAAC	3.2E-1	3.6E-1	1.0E+0	1.0E+0	9.0E-1	3.6E-1	8.0E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
ProtVec	1.9E-1	3.6E-1	1.0E+0	1.0E+0	6.8E-1	4.2E-1	1.3E-2	2.9E-3	5.4E-3	5.4E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
AAC	7.8E-2	1.0E-1	9.0E-1	6.8E-1	1.0E+0	1.0E+0	2.7E-1	2.5E-2	1.3E-2	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
Mut2Vec^	2.9E-3	2.9E-3	3.6E-1	4.2E-1	1.0E+0	1.0E+0	2.9E-3	2.9E-3	2.9E-3	5.4E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
Learned-Vec	2.9E-3	2.9E-3	8.0E-3	1.3E-2	2.7E-1	2.9E-3	1.0E+0	1.6E-1	8.0E-3	2.5E-2	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
CPCProt	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.5E-2	2.9E-3	1.6E-1	1.0E+0	2.5E-2	4.5E-2	5.4E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
K-Sep	2.9E-3	2.9E-3	2.9E-3	5.4E-3	1.3E-2	2.9E-3	8.0E-3	2.5E-2	1.0E+0	9.0E-1	3.6E-1	2.3E-1	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
MSA-Transformer	2.9E-3	2.9E-3	2.9E-3	5.4E-3	2.9E-3	5.4E-3	2.5E-2	4.5E-2	9.0E-1	1.0E+0	7.6E-1	2.7E-1	2.5E-2	1.3E-2	8.0E-3	4.5E-2	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
UniRep	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	5.4E-3	3.6E-1	7.6E-1	1.0E+0	5.4E-1	5.4E-3	2.9E-3	5.4E-3	1.3E-2	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
TAPE-BERT-PFAM^	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.3E-1	2.7E-1	5.4E-1	1.0E+0	4.5E-2	5.4E-3	1.8E-2	4.5E-2	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
BLAST	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.5E-2	5.4E-3	4.5E-2	1.0E+0	1.8E-2	1.8E-2	1.6E-1	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
HMMER	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	1.3E-2	2.9E-3	5.4E-3	1.8E-2	1.0E+0	9.8E-1	5.4E-1	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
ProtBERT-BFD	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	8.0E-3	5.4E-3	1.8E-2	1.8E-2	9.8E-1	1.0E+0	6.8E-1	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3
ProtXLNet	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	4.5E-2	1.3E-2	4.5E-2	1.6E-1	5.4E-1	6.8E-1	1.0E+0	1.6E-1	3.4E-2	2.9E-3	5.4E-3	2.9E-3
SeqVec	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	1.6E-1	1.0E+0	7.8E-2	1.3E-2	5.4E-3	2.9E-3
PFAM^	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	3.4E-2	7.8E-2	1.0E+0	1.8E-2	2.9E-3	2.9E-3
ESM-1b	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	1.3E-2	1.8E-2	1.0E+0	8.3E-1	8.3E-1
ProtALBERT	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	5.4E-3	5.4E-3	2.9E-3	8.3E-1	1.0E+0	4.9E-1
ProtT5-XL	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	2.9E-3	8.3E-1	4.9E-1	1.0E+0

(d)

	APAAC	Gene2Vec^	TCGA-Embedding	ProtVec	AAC	Mut2Vec^	Learned-Vec	UniRep	MSA-Transformer	K-Sep	CPCProt	BLAST	ProtXLNet	TAPE-BERT-PFAM^	HMMER	ProtBERT-BFD	PFAM^	SeqVec	ESM-1b	ProtALBERT	ProtT5-XL
APAAC	1.0E+0	3.7E-2	5.0E-2	1.9E-2	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
Gene2Vec^	3.7E-2	1.0E+0	6.2E-1	8.2E-2	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
TCGA-Embedding	5.0E-2	6.2E-1	1.0E+0	1.0E-1	1.9E-2	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
ProtVec	1.9E-2	8.2E-2	1.0E-1	1.0E+0	1.3E-1	1.0E-1	3.5E-3	2.7E-2	8.9E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
AAC	3.5E-3	3.5E-3	1.9E-2	1.3E-1	1.0E+0	9.7E-1	6.2E-3	3.7E-2	1.4E-2	1.4E-2	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
Mut2Vec^	6.2E-3	3.5E-3	3.5E-3	1.0E-1	9.7E-1	1.0E+0	1.4E-2	6.4E-2	1.4E-2	8.9E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
Learned-Vec	3.5E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	1.4E-2	1.0E+0	1.5E-1	1.3E-1	8.2E-2	8.9E-3	3.5E-3	1.4E-2	8.9E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
UniRep	3.5E-3	6.2E-3	6.2E-3	2.7E-2	3.7E-2	6.4E-2	1.5E-1	1.0E+0	9.7E-1	9.2E-1	6.2E-1	9.2E-1	6.9E-1	8.2E-2	6.4E-2	3.5E-3	3.5E-3	3.5E-3	8.9E-3	3.5E-3	3.5E-3
MSA-Transformer	3.5E-3	3.5E-3	3.5E-3	8.9E-3	1.4E-2	1.4E-2	1.3E-1	9.7E-1	1.0E+0	9.7E-1	9.7E-1	6.9E-1	1.3E-1	8.2E-2	1.0E-1	3.5E-3	1.4E-2	3.5E-3	3.5E-3	3.5E-3	3.5E-3
K-Sep	6.2E-3	3.5E-3	3.5E-3	6.2E-3	1.4E-2	8.9E-3	8.2E-2	9.2E-1	9.7E-1	1.0E+0	9.7E-1	6.9E-1	3.2E-1	6.4E-2	8.9E-3	6.2E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
CPCProt	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	8.9E-3	6.2E-1	9.7E-1	9.7E-1	1.0E+0	1.3E-1	1.5E-1	1.0E-1	3.5E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3
BLAST	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	9.2E-1	6.9E-1	6.9E-1	1.3E-1	1.0E+0	3.7E-1	2.7E-1	2.7E-2	6.2E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3
ProtXLNet	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	1.4E-2	6.9E-1	1.3E-1	3.2E-1	1.5E-1	3.7E-1	1.0E+0	1.0E+0	9.7E-1	1.5E-1	2.7E-1	1.0E-1	3.5E-3	6.2E-3	3.5E-3
TAPE-BERT-PFAM^	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	6.2E-3	8.9E-3	8.2E-2	8.2E-2	6.4E-2	1.0E-1	2.7E-1	1.0E+0	1.0E+0	4.9E-1	6.4E-2	1.3E-1	3.5E-3	3.7E-2	3.5E-3	3.5E-3
HMMER	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	6.4E-2	1.0E-1	8.9E-3	3.5E-3	2.7E-2	9.7E-1	4.9E-1	1.0E+0	6.2E-3	5.0E-2	3.5E-3	6.2E-3	3.5E-3	3.5E-3
ProtBERT-BFD	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	6.2E-3	1.5E-1	6.4E-2	6.2E-3	1.0E+0	6.2E-1	2.7E-2	8.2E-2	3.5E-3	3.5E-3
PFAM^	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	1.4E-2	6.2E-3	3.5E-3	6.2E-3	2.7E-1	1.3E-1	6.4E-2	6.2E-1	1.0E+0	1.5E-1	6.4E-2	3.5E-3	3.5E-3
SeqVec	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	1.0E-1	3.5E-3	3.5E-3	2.7E-2	1.5E-1	1.0E+0	1.9E-1	1.9E-2	3.5E-3
ESM-1b	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	8.9E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.7E-2	6.2E-3	8.2E-2	6.4E-2	1.9E-1	1.0E+0	6.4E-1	5.0E-2
ProtALBERT	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	6.2E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	1.9E-2	6.4E-1	1.0E+0	1.4E-2
ProtT5-XL	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3	5.0E-2	1.4E-2	1.0E+0

Table S9. Protein-protein binding affinity estimation benchmark performance results. Mean squared error (MSE), mean absolute error (MAE) were calculated by taking the mean of 10-fold cross validation results. All metrics are multiplied with 10^2 for easy comprehension. Methods signed with "†" displayed better performance compared to the best method in the PIPR study [140] study.

	Method name	Performance results*		
		MSE x 10^2	**MAE x 10^2**	**Correlation x 10^2**
Methods from our study	K-Sep	0.97	7.21	76.13
	APAAC	1.79	10.23	47.72
	PFAM	2.26	11.69	19.17
	AAC	1.85	10.59	46.05
	ProtVec	1.13	8.09	71.91
	Learned-Vec	1.18	8.13	70.34
	SeqVec†	0.53	5.24	88.01
	CPCProt	0.73	6.26	83.09
	MSA-Transformer	0.908016	7.09502	78.35911
	ProtBERT-BFD	0.57	5.21	87.40
	TAPE-BERT-PFAM	0.57	5.57	87.14
	ESM-1b†	0.48	5.00	89.15
	ProtALBERT†	**0.42**	**4.57**	**90.71**
	ProtXLNet	0.61	5.70	86.15
	UniRep	0.73	6.39	82.99
	ProtT5-XL	0.60	5.46	86.66
Methods from the PIPR study [140]	AC*	1.70	9.56	56.4
	CTD*	1.86	10.2	50.1
	SCNN*	0.87	6.49	83.1
	SRGRU*	0.95	7.08	81.2
	SRRCNN (PIPR)*	0.63	5.48	87.3

*** MSE:** mean squared error, **MAE:** mean absolute error, **Corr:** Pearson correlation, **AC**: Autocovariance [168], **CTD**: Composition- Transition-Distribution [169], **SCNN**: Siamese Convolutional Neural Network, **SRGRU**: Siamese Residual Gated Recurrent Unit **PIPR**: Protein–Protein Interaction Prediction Based on Siamese Residual Recurrent CNN.

Table S10. Statistical significance test between representation methods used in the protein-protein binding affinity estimation task. FDR values were calculated using the Wilcoxon Rank Sum test and Benjamini-Hochberg multiple test correction. FDR values are based on results of **(a)** Mean squared error, **(b)** Mean absolute error scores. Color codes represent FDR values (pink: FDR > 0.05, yellow: 0.05 > FDR > 0.0005, green: FDR =< 0.0005).

(a)

	ProtALBERT	ESM-1b	SeqVec	ProtBERT-BFD	ProtT5-XL	ProtXLNet	CPCPROT	UniRep	MSA_Transformer	KSEP	ProtVec	Learned-Vec	APAAC	AAC	PFAM	TAPE-BERT-PFAM
ProtALBERT	1.0E+0	3.3E-2	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
ESM-1b	3.3E-2	1.0E+0	3.3E-2	5.7E-2	2.4E-2	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
SeqVec	2.5E-3	3.3E-2	1.0E+0	8.3E-1	2.6E-1	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
ProtBERT-BFD	2.5E-3	5.7E-2	8.3E-1	1.0E+0	5.4E-1	1.2E-1	4.9E-3	4.4E-2	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
ProtT5-XL	2.5E-3	2.4E-2	2.6E-1	5.4E-1	1.0E+0	4.8E-1	5.7E-2	7.4E-2	5.1E-3	4.9E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
ProtXLNet	2.5E-3	2.5E-3	2.5E-3	1.2E-1	4.8E-1	1.0E+0	2.5E-3	2.5E-3	5.1E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
CPCPROT	2.5E-3	2.5E-3	2.5E-3	4.9E-3	5.7E-2	2.5E-3	1.0E+0	9.1E-1	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
UniRep	2.5E-3	2.5E-3	2.5E-3	4.4E-2	7.4E-2	2.5E-3	9.1E-1	1.0E+0	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
MSA_Transformer	2.6E-3	2.6E-3	2.6E-3	2.6E-3	5.1E-3	5.1E-3	2.6E-3	2.6E-3	2.6E-3	1.0E+0	2.6E-1	2.6E-3	2.6E-3	2.6E-3	2.6E-3	2.6E-3
KSEP	2.5E-3	2.5E-3	2.5E-3	2.5E-3	4.9E-3	2.5E-3	2.5E-3	2.5E-3	1.0E+0	1.0E+0	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3
ProtVec	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-1	2.5E-3	1.0E+0	7.4E-2	2.5E-3	2.5E-3	2.5E-3	2.5E-3
Learned-Vec	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	7.4E-2	1.0E+0	2.5E-3	2.5E-3	2.5E-3	2.5E-3
APAAC	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	1.0E+0	2.2E-1	2.5E-3	2.5E-3
AAC	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.2E-1	1.0E+0	2.5E-3	2.5E-3
PFAM	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	1.0E+0	2.5E-3
TAPE-BERT-PFAM	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.6E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	2.5E-3	1.0E+0

(b)

	ProtALBERT	ESM-1b	ProtBERT-BFD	SeqVec	ProtT5-XL	ProtXLNet	CPCPROT	UniRep	MSA_Transformer	KSEP	ProtVec	Learned-Vec	APAAC	AAC	PFAM	TAPE-BERT-PFAM
ProtALBERT	1.0E+0	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
ESM-1b	2.3E-3	1.0E+0	4.1E-2	3.1E-2	4.6E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
ProtBERT-BFD	2.3E-3	4.1E-2	1.0E+0	8.2E-1	4.1E-2	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
SeqVec	2.3E-3	3.1E-2	8.2E-1	1.0E+0	2.2E-2	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
ProtT5-XL	2.3E-3	4.6E-3	4.1E-2	2.2E-2	1.0E+0	1.6E-2	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
ProtXLNet	2.3E-3	2.3E-3	2.3E-3	2.3E-3	1.6E-2	1.0E+0	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
CPCPROT	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	1.0E+0	9.2E-2	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
UniRep	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	9.2E-2	1.0E+0	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
MSA_Transformer	2.4E-3	2.4E-3	2.4E-3	2.4E-3	2.4E-3	2.4E-3	2.4E-3	2.4E-3	2.4E-3	1.0E+0	6.7E-1	2.4E-3	2.4E-3	2.4E-3	2.4E-3	2.4E-3
KSEP	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	1.0E+0	1.0E+0	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3
ProtVec	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	6.7E-1	2.3E-3	1.0E+0	6.8E-1	2.3E-3	2.3E-3	2.3E-3	2.3E-3
Learned-Vec	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	6.8E-1	1.0E+0	2.3E-3	2.3E-3	2.3E-3	2.3E-3
APAAC	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	1.0E+0	2.2E-2	2.3E-3	2.3E-3
AAC	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.2E-2	1.0E+0	2.3E-3	2.3E-3
PFAM	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	1.0E+0	2.3E-3
TAPE-BERT-PFAM	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.4E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	2.3E-3	1.0E+0

Table S11. Number samples in train-test folds of the random split, Uniclust50, Uniclust30 and MMSEQ-15 datasets in the drug target protein family classification benchmark.

Random split dataset										
Folds	**Train**					**Test**				
	Enzyme	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**	**Enzyme**	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**
Fold1	3919	752	926	315	919	441	79	108	32	100
Fold2	3936	737	926	302	930	424	94	108	45	89
Fold3	3914	751	936	310	920	446	80	98	37	99
Fold4	3924	751	920	308	928	436	80	114	39	91
Fold5	3925	737	944	314	911	435	94	90	33	108
Fold6	3933	748	919	316	915	427	83	115	31	104
Fold7	3916	746	929	321	919	444	85	105	26	100
Fold8	3905	774	939	302	911	455	57	95	45	108
Fold9	3928	742	929	318	914	432	89	105	29	105
Fold10	3930	741	938	317	905	430	90	96	30	114
50% Similarity threshold (Uniclust50)										
Folds	**Train**					**Test**				
	Enzyme	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**	**Enzyme**	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**
Fold1	3822	676	910	309	876	441	79	108	32	100
Fold2	3839	668	906	286	894	424	94	108	45	89
Fold3	3807	685	899	301	901	446	80	98	37	99
Fold4	3830	671	901	297	894	436	80	114	39	91
Fold5	3822	641	923	307	900	435	94	90	33	108
Fold6	3857	669	899	307	861	427	83	115	31	104
Fold7	3833	653	910	320	877	444	85	105	26	100
Fold8	3796	715	928	291	863	455	57	95	45	108
Fold9	3837	657	901	312	886	432	89	105	29	105
Fold10	3851	654	913	313	862	430	90	96	30	114
30% Similarity threshold (Uniclust30)										
Folds	**Train**					**Test**				
	Enzyme	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**	**Enzyme**	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**
Fold1	3750	522	890	301	845	441	79	108	32	100
Fold2	3780	496	855	274	903	424	94	108	45	89
Fold3	3738	524	863	279	904	446	80	98	37	99
Fold4	3767	530	851	287	873	436	80	114	39	91
Fold5	3754	512	895	305	842	435	94	90	33	108
Fold6	3797	571	852	304	784	427	83	115	31	104
Fold7	3771	479	875	317	866	444	85	105	26	100
Fold8	3752	573	897	272	814	455	57	95	45	108
Fold9	3758	506	886	305	853	432	89	105	29	105
Fold10	3807	488	879	300	834	430	90	96	30	114

15% Similarity threshold (MMSEQ-15)										
Folds	**Train**					**Test**				
	Enzyme	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**	**Enzyme**	**Membrane Receptor**	**Transcription Factor**	**Ion Channel**	**Others**
Fold1	3139	198	577	222	710	441	79	108	32	100
Fold2	3126	182	574	169	795	424	94	108	45	89
Fold3	3039	213	591	189	814	446	80	98	37	99
Fold4	3150	225	567	201	703	436	80	114	39	91
Fold5	3060	198	667	224	697	435	94	90	33	108
Fold6	3135	207	588	245	671	427	83	115	31	104
Fold7	3103	196	584	264	699	444	85	105	26	100
Fold8	3062	265	622	178	719	455	57	95	45	108
Fold9	3149	205	597	245	650	432	89	105	29	105
Fold10	3165	201	640	226	614	430	90	96	30	114