Supplemental Information

Empirical Comparison and Analysis of Web-Based

DNA N^4-Methylcytosine Site Prediction Tools

Balachandran Manavalan, Md. Mehedi Hasan, Shaherin Basith, Vijayakumar Gosu, Tae-Hwan Shin, and Gwang Lee

Table S1. Performance of existing predictors on validation set.

Species	Methods	MCC	BACC	Sn	Sp	AUC	TP	TN	FN	FP	P-value
Geob. pickeringii	iDNA4mC	0.118	0.610	0.532	0.687	0.653	2364	41251	2081	18754	—
	4mCPred_I	0.072	0.571	0.620	0.523	0.598	2756	31363	1689	28642	<0.00001
	4mCPred_II	0.069	0.568	0.620	0.516	0.589	2756	30981	1689	29024	<0.00001
	4mCCNN	-0.008	0.492	0.507	0.478	0.492	2254	28671	2191	31334	<0.00001
	4mcPred-SVM	-0.014	0.486	0.557	0.415	0.485	2477	24932	1968	35073	<0.00001
	DNA4mC-LIP	-0.022	0.479	0.582	0.376	0.505	2585	22544	1860	37461	<0.00001
	Meta-4mCpred	-0.064	0.441	0.560	0.321	0.430	2491	19256	1954	40749	<0.00001
E. coli	4mCCNN	0.110	0.793	0.781	0.805	0.873	523	95232	147	23034	—
	DNA4mC-LIP	0.102	0.788	0.804	0.771	0.866	539	91181	131	27085	0.577254
	4mCPred_II	0.096	0.788	0.854	0.722	0.865	572	85404	98	32862	0.524711
	4mcPred-SVM	0.105	0.782	0.766	0.799	0.858	513	94501	157	23765	0.237642
	Meta-4mCpred	0.094	0.771	0.791	0.751	0.842	530	88854	140	29412	0.016821
	4mCPred_I	0.097	0.768	0.757	0.780	0.842	507	92264	163	26002	0.016821
	iDNA4mC	0.056	0.668	0.622	0.714	0.728	417	84467	253	33799	<0.00001
C. elegans	DNA4mC-LIP	0.478	0.825	0.832	0.817	0.900	16049	118144	3240	26409	—
	4mCCNN	0.486	0.824	0.815	0.832	0.904	15730	120310	3559	24243	0.58093
	4mCPred_II	0.453	0.820	0.859	0.781	0.898	16567	112949	2722	31604	0.349634
	4mCPred_I	0.474	0.819	0.815	0.824	0.892	15721	119041	3568	25512	0.00022
	4mcPred-SVM	0.405	0.781	0.770	0.791	0.858	14851	114355	4438	30198	<0.00001
	iDNA4mC	0.401	0.775	0.751	0.799	0.854	14491	115436	4798	29117	<0.00001
	Meta-4mCpred	0.376	0.763	0.750	0.777	0.822	14459	112343	4830	32210	<0.00001
A. thaliana	DNA4mC-LIP	0.561	0.796	0.778	0.813	0.868	57432	145362	16353	33349	—
	4mCPred_I	0.528	0.780	0.767	0.794	0.852	56574	141902	17211	36809	<0.00001
	4mCPred_II	0.520	0.776	0.758	0.794	0.849	55924	141813	17861	36898	<0.00001
	4mCCNN	0.513	0.772	0.756	0.788	0.844	55815	140847	17970	37864	<0.00001
	Meta-4mCpred	0.515	0.771	0.740	0.803	0.844	54581	143445	19204	35266	<0.00001
	4mcPred-SVM	0.509	0.770	0.754	0.786	0.843	55658	140554	18127	38157	<0.00001
	iDNA4mC	0.220	0.609	0.439	0.779	0.672	32416	139291	41369	39420	<0.00001
D. melanogaster	4mCCNN	0.515	0.822	0.885	0.758	0.895	40549	159036	5260	50654	—
	DNA4mC-LIP	0.498	0.817	0.913	0.721	0.896	41818	151282	3991	58408	0.481122
	4mCPred_I	0.490	0.809	0.883	0.735	0.885	40450	154175	5359	55515	<0.00001
	Meta-4mCpred	0.477	0.803	0.884	0.722	0.862	40479	151326	5330	58364	<0.00001
	4mCPred_II	0.449	0.790	0.908	0.672	0.864	41585	140817	4224	68873	<0.00001
	4mcPred-SVM	0.463	0.793	0.868	0.719	0.867	39748	150831	6061	58859	<0.00001
	iDNA4mC	0.238	0.643	0.542	0.745	0.712	24819	156122	20990	53568	<0.00001
Geoa. subterraneus	iDNA4mC	0.150	0.575	0.449	0.701	0.611	7885	7999	9688	3405	—
	4mCPred_II	0.097	0.549	0.604	0.495	0.574	10612	5642	6961	5762	<0.00001
	4mCCNN	0.087	0.545	0.553	0.536	0.555	9724	6109	7849	5295	<0.00001
	DNA4mC-LIP	0.076	0.539	0.568	0.510	0.551	9978	5814	7595	5590	<0.00001
	4mCPred_I	0.070	0.536	0.591	0.481	0.557	10383	5482	7190	5922	<0.00001
	4mcPred-SVM	0.069	0.535	0.571	0.500	0.546	10031	5702	7542	5702	<0.00001
	Meta-4mCpred	0.043	0.522	0.574	0.469	0.524	10094	5352	7479	6052	<0.00001
M. musculus	i4mC-Mouse	0.020	0.578	0.808	0.348	0.633	761	86181	181	161642	—
	4mCpred-EL	0.018	0.571	0.773	0.370	0.612	728	91703	214	156120	0.129443
F. vesca	iDNA-MS	0.417	0.847	0.876	0.818	0.930	18710	238038	2640	52819	—
	i4mC-ROSE	0.386	0.802	0.758	0.847	0.889	16179	246439	5171	44417	<0.00001

The first and second column respectively represent the species and method names. Methods are ranked according to its BACC for each species. The top method AUC value is compared with other methods and computed the *P*-value using two-tailed test. MCC: Matthews correlation coefficient; ACC: accuracy; Sn: sensitivity; Sp: specificity. If the value is not provided in the literature, it is mentioned as '—'.