

## Supplementary materials:

### **PLMLA: Prediction of lysine methylation and lysine acetylation by combining multiple features**

Shao-Ping Shi<sup>a,b</sup>, Jian-Ding Qiu<sup>a\*</sup>, Xing-Yu Sun<sup>a</sup>, Sheng-Bao Suo<sup>a</sup>, Shu-Yun Huang<sup>a</sup>,

Ru-Ping Liang<sup>a</sup>

<sup>a</sup> Department of Chemistry, Nanchang University, Nanchang 330031, China

<sup>b</sup> Department of mathematics, Nanchang University, Nanchang 330031, China

***Running title: Prediction of lysine methylation and acetylation***

\* Corresponding Author: Jian-Ding Qiu, Department of Chemistry, Nanchang

University, Nanchang 330031, China. Tel: + 86 791 3969518, e-mail:

jdqiu@ncu.edu.cn

## Training data sets:

**546 experimental methyllysine sites from 157 proteins.(txt)**

**2842 experimental acetyllysine sites from 1639 proteins.(txt)**

**Table S1. The 16 test proteins covering 33 experimental methyllysine.**

Protein name	Residue position	Flanking residues
Q63598 PLST_RAT	447	DWSKVNKPPYPKL
P61289 PSME3_HUMAN	121	LVDIIEKVKPEIR
	212	VTEIDEKEYISLR
Q5XI73 GDIR1_RAT	50	DDESLRKYKEALL
	52	ESLRKYKEALLGR
Q8CBD1 NRIP1_MOUSE	591	CSTPASKLTNTAP
	653	EEQRCKQLLSGN
	778	PLNHDAKSAPFLG
Q14683 SMC1A_HUMAN	149	IAMKNPKERTALF
	1133	NLSGGEKTVAALA
Q9UQE7 SMC3_HUMAN	245	LDELSAKRETSGE
	409	DQAINDKKRQIAA
	427	EDTEANKEKNLEQ
	429	TEANKEKNLEQYN
	445	QDLNEVKARVEEL
Q8WVM7 STAG1_HUMAN	486	QQALAAKREDLEK
	997	VNFSEQKEKLIKR
	333	LNDSYLKYVGWTL
P17948 VGFR1_HUMAN	618	HLDALLKQIKFVV
	831	ERLKLGKSLGRGA
P31000 VIME_RAT	143	QLKGQGKSRLGDL
P05712 RAB2A_RAT	8	AYAYLFKYIIIGD
P0CH08 RL401_YEAST	98	DKSVCRKCYARLP
P09543 CN37_HUMAN	27	MSSSGAKDKPELQ
	29	SSGAKDKPELQFP
P81543 FERA_SULME	29	PVENPKVLGIHG
P01241 SOMA_HUMAN	96	REETQQKSNLELL
P05496 AT5G1_HUMAN	104	ARNPSLKQQLFSY
	120	NIAQQAKFFNFTV
	130	FTVAAGKILNITG
	156	AQNALTKVHGGAA
	204	TLGANSKIVNGGN
Q53020 OMP_B_RICPR	231	NTFAGIKTINIDD

**Table S2. The 30 test proteins covering 36 experimental acetyllysine.**

Protein name	Residue position	Flanking residues
P30153 2AAA_HUMAN	280	VGPEITKTDLVPA
	42	FRTFGRKHEEELS
Q61694 3BHS5_MOUSE	60	AKVRVLKGDILDA
	362	WIGTLVKQHRETL
Q9NRA8 4ET_HUMAN	486	DMTAFNKLVTSMK
Q969T7 5NT3L_HUMAN	248	IGFLNDKVEERRE
P56378 68MP_HUMAN	14	NIWIPMKPYITKV
P52697 6PGL_ECOLI	287	NVDHSGKYLIAAG
O95336 6PGL_HUMAN	180	PISDSPKPPPQRV
Q8IZ83 A16A1_HUMAN	50	VNGKWLKPEHRNS
P54619 AAKG1_HUMAN	264	LDVSVTKALQHRS
P54646 AAPK2_HUMAN	69	REIQNLKLFRRPH
Q9UDR5 AASS_HUMAN	707	PNRDSTKYAEIYG
Q9QXD1 ACOX2_MOUSE	667	QENPAYKKYIRPL
O14561 ACPM_HUMAN	92	VLKLYDKIDPEKL
P0A9Q7 ADHE_ECOLI	358	AVEKAEKLVAMGG
P11766 ADHX_HUMAN	366	ELMHSGKSIRTVV
Q9H2P0 ADNP_HUMAN	1035	KNSSYGKVEGFWS
	1042	VEGFWSKDQSQWK
Q16186 ADRM1_HUMAN	21	SRGASNKYLVEFR
Q9H0C2 ADT4_HUMAN	104	ALNFAFKDKYKQL
P51825 AFF1_HUMAN	681	SLPAPSKALSGPE
Q9UHB7 AFF4_HUMAN	79	KLVAIPKPTVPPS
Q8JZQ2 AFG32_MOUSE	307	EIDVKFKDVAGCE
	542	SDAINEKHFEQAI
Q9ATB4 TAD2B_ARATH	216	DRSFGGKKPVSTS
P09204 TBA1_CHLRE	40	GQMPSDKTIGGGD
P0ADN2 YIFE_ECOLI	15	NRYFDNKHYPRGF
P0AF96 YJGK_ECOLI	36	AETPKGKHDIEN
P0AF36 ZAPB_ECOLI	10	SLEVFEEKLEAKVQ
O43298 ZBT43_HUMAN	247	KRWIHKVPERLEQ
Q921H8 THIKA_MOUSE	265	KLKPAFKDGGSTT
	395	TLLNELKRRGRRA
Q8NFI4 F10A5_HUMAN	153	LAILYAKRASVVF
P41212 ETV6_HUMAN	11	PAQCSIKQERISY
	302	GLHREGKPINLSH

**Table S3. The predictive accuracy of different sub-sequence numbers (*J*) by using the EBGW feature (%).**

The number of sub-sequences	3	4	5	6
Methyllysine	70.76	77.58	78.65	76.72
Acetyllysine	67.06	66.27	69.93	69.03

**Table S4. The position weight matrix of residues surrounding 546 methyllysine sites, where the red digital is the maximum occurrence frequency of twenty amino acids at each position.**

Amino Acids	The occurrence frequency of twenty amino acids at each position												
	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
A	44	33	33	37	49	38	0	37	34	35	42	39	35
C	5	4	4	5	6	6	0	6	4	10	4	4	5
D	44	34	31	42	27	32	0	29	31	31	27	20	31
E	50	51	55	52	52	50	0	33	36	54	56	46	35
F	27	23	20	26	17	34	0	9	28	27	23	27	25
G	33	33	38	37	39	42	0	38	54	48	38	44	49
H	7	8	9	10	12	5	0	6	10	10	12	12	5
I	34	33	32	32	35	34	0	34	37	34	41	26	37
K	42	32	34	39	32	40	546	47	31	46	34	41	46
L	37	61	51	42	47	58	0	46	59	47	39	43	50
M	15	16	9	7	13	12	0	16	10	7	7	12	13
N	27	33	24	30	26	19	0	23	18	23	20	16	26
P	25	20	21	25	19	23	0	26	28	19	23	17	19
Q	19	23	16	14	20	20	0	23	16	19	18	18	19
R	27	21	21	25	26	39	0	29	27	19	30	32	24
S	40	34	41	47	35	30	0	30	23	36	39	51	31
T	23	21	27	25	35	13	0	33	30	28	30	27	32
V	34	37	51	31	32	32	0	50	39	34	36	47	43
W	3	4	3	4	6	2	0	8	6	5	5	5	2
Y	10	25	26	16	18	17	0	23	25	14	22	19	18

**Table S5. The position weight matrix of residues surrounding 2842 acetyllysine sites, where the red digital is the maximum occurrence frequency of twenty amino acids at each position.**

Amino Acids	The occurrence frequency of twenty amino acids at each position												
	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
A	208	223	204	224	214	232	0	167	177	169	203	210	206
C	28	27	34	29	22	17	0	6	28	18	24	32	25
D	127	123	150	134	139	153	0	172	123	210	146	116	150
E	190	147	247	228	145	317	0	231	270	241	219	192	198
F	115	106	75	149	287	100	0	178	164	108	122	116	89
G	173	165	205	183	187	323	0	161	164	185	200	182	164
H	61	67	86	74	61	59	0	133	40	69	63	66	60
I	137	145	125	154	208	139	0	139	237	107	150	134	138
K	316	313	255	183	117	66	2842	68	157	308	308	305	324
L	259	267	202	252	342	298	0	285	381	211	252	291	271
M	60	55	45	50	49	63	0	58	58	42	52	49	54
N	133	118	129	125	107	116	0	111	70	116	86	110	108
P	125	127	150	121	128	84	0	202	136	143	117	115	131
Q	113	125	120	147	92	142	0	78	83	123	125	124	128
R	217	222	168	111	68	38	0	125	148	252	215	215	213
S	179	179	215	196	143	225	0	129	146	176	168	157	183
T	138	127	147	162	102	143	0	122	107	138	151	144	133
V	153	195	182	194	199	193	0	139	202	129	140	174	168
W	17	23	20	24	66	38	0	45	32	17	27	25	19
Y	93	88	83	102	166	96	0	293	119	80	74	85	80

**Table S6. The position weight matrix of residues surrounding 2842 non-methyllysine and non-acetyllysine, where the red digital is the maximum occurrence frequency of twenty amino acids at each position.**

Amino Acids	The occurrence frequency of twenty amino acids at each position												
	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
A	234	193	210	168	231	172	0	215	190	176	190	206	184
C	29	52	25	45	34	26	0	40	44	40	43	35	32
D	146	149	190	152	149	139	0	163	126	164	153	122	157
E	228	231	284	262	228	268	0	253	212	250	260	268	211
F	108	104	78	83	116	104	0	84	118	106	103	97	112
G	196	172	144	194	192	211	0	157	188	213	195	226	215
H	77	75	58	82	66	59	0	59	64	51	56	67	79
I	156	151	158	113	133	130	0	143	149	141	153	135	171
K	209	243	226	235	217	223	2842	212	215	217	230	235	224
L	237	283	245	285	269	286	0	247	309	248	228	227	270
M	54	59	43	54	58	63	0	64	72	57	64	62	55
N	91	85	92	120	127	102	0	105	92	95	107	98	96
P	155	153	158	163	159	141	0	180	127	154	160	188	169
Q	120	124	141	128	93	146	0	143	123	112	148	111	117
R	137	144	159	128	158	183	0	156	173	136	132	152	131
S	201	207	202	219	191	204	0	188	195	257	224	212	222
T	160	138	163	144	155	113	0	108	133	160	126	120	146
V	172	174	161	147	190	158	0	206	192	168	175	179	159
W	40	23	26	28	20	30	0	28	33	35	23	24	25
Y	92	82	79	92	56	84	0	91	87	62	72	78	67

**Table S7. Performance comparison between PLMLA and MASA on the dataset adopted in MASA method.**

Model type	Prediction methods	The predictive performance (%)			
		Sn	Sp	Acc	MCC
Methyllysine	MASA	75.1	74.0	74.6	56.1
Methyllysine	PLMLA	76.7	81.6	79.1	58.3

**Table S8. Performance comparison between PLMLA and N-Ace on the dataset adopted in N-Ace method.**

Model type	Prediction methods	The predictive performance (%)			
		Sn	Sp	Acc	MCC
Acetyllysine	N-Ace	63.5	86.0	74.9	51.0
Acetyllysine	PLMLA	88.5	85.6	87.0	74.1