Supplementary Materials:

SUMOsp: a web server for sumoylation sites prediction

Yu Xue[1,†], Fengfeng Zhou[2,†], Chuanhai Fu[1], Ying Xu[2,*] and Xuebiao Yao[1,3,*]

[1]Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Sciences, and the University of Science and Technology of China, Hefei, China 230027;

[2]Computational Systems Biology Laboratory, Department of Biochemical and Molecular Biology, and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

[3]Department of Physiology and Cancer Research Program, Morehouse School of Medicine, Atlanta, GA 30310; USA

Running title: sumoylation sites prediction

*To whom correspondence should be addressed.

Ying Xu: Tel: 706-542-9779; Email: xyn@bmb.uga.edu

Xuebiao Yao: Tel: 86-551-3606304; Email: yaoxb@ustc.edu.cn

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Supplementary Tables

Table S1. Sumoylation sites which don't match the consensus motif -K-X-E.

(A) There are about 23% (56/239) real sumoylation sites which are non-canonical sites. The PMIDs for each curated sumoylation sites are also shown.

(B) The prediction results of SUMOsp on these sites

Protein Name	Swissprot ID	Organism	Sumoylation Site	Motif	PMID
Rep78	O56136	Virus	K84	E**K**GE	15527853
CENP-C	Q03188	Human	K721	P**K**NR	15272016
			K746	L**K**PL	15272016
Ubc9	P63279	Human	K153	A**K**KF	15272016
PCNA	P15873	Yeast	K164	T**K**ET	12226657;15121847;15542864
TEL/ETV6	P41212	Human	K99	T**K**ED	12626745;15107848
Huntingtin	P42858	Human	K6	E**K**LM	15064418
			K9	M**K**AF	15064418
			K15	L**K**SF	15064418
Smad4	Q13485	Human	K113	V**K**YC	12621041;12740389;15028714
Bach2	Q9BYV9	Human	K202	E**K**EE	15060166
			K421	C**K**QE	15060166
NEMO	Q9Y6K9	Human	K309	Y**K**AD	14651848
Dlp/Drp	O00429	Human	K38	G**K**SS	14972687
Topors	Q9UNR9	Human	K560	K**K**EE	14516784
Fak1	Q05397	Human	K152	V**K**SD	14500712
Cdc3	P32457	Yeast	K11	I**K**QD	12149243;12761287
			K287	A**K**SD	12761287
			K465	Q**K**SE	12761287
DJ-1	Q99497	Human	K130	A**K**DK	12761214
CREB1	P16220	Human	K285	R**K**RE	12552083
			K304	K**K**KE	12552083
Bace	P56817	Human	K275	L**K**MD	12506199
Axin-1	O15169	Human	K857	G**K**VE	12223491
			K860	E**K**VD	12223491
Daxx	Q9UER7	Human	K630	C**K**KS	12150977
			K631	K**K**SR	12150977
NCoA-2	Q15596	Human	K788	E**K**EE	12060666
PML	P29590	Human	K65	A**K**CP	9756909;12149243
HIPK-2	Q9H2X6	Human	K1191	A**K**VN	12149243

VE1	P03116	Virus	K514	I**K**AP	11005821;12149243
LAF1	Q9M0K4	Arabidopsis	K258	K**K**QE	11581165
BZLF1	P03206	Virus	K12	V**K**FT	11160742
Rta	P03209	Virus	K19	I**K**KQ	15229220
			K213	S**K**TG	15229220
			K517	V**K**AL	15229220
HIP2	P61086	Human	K13	F**K**EV	15723079
GAG	P04591	Virus	K474	Q**K**QE	15613319
SOD1	P00445	Yeast	K69	K**K**TH	15596868
RSC58	Q07979	Yeast	K322	E**K**NE	15596868
RPL25	P04456	Yeast	K60	Y**K**VI	15542864
SIR4	P11978	Yeast	K1128	V**K**NV	15542864
YMR192W	Q04322	Yeast	K498	L**K**MG	15542864
RB1	P06400	Human	K720	L**K**FK	15806172
KCNK1	O00180	Human	K274	L**K**KF	15820677
ATXN1	P54253	Human	K16	K**K**RE	15824120
			K194	H**K**AE	15824120
			K610	L**K**ID	15824120
			K697	V**K**KG	15824120
			K746	L**K**FP	15824120
BLM	P54132	Human	K317	S**K**CL	15829507
			K331	R**K**ED	15829507
			K344	S**K**PE	15829507
			K347	E**K**MS	15829507
CASP2	P42575	Human	K77	A**K**VG	15882978
NFE2	Q16621	Human	K368	T**K**ME	16287851

Table S1 (A)

Swissprot ID	Sumoylation Site	GPS Score	Motif-X
O56136	K84	1.08	1
Q03188	K721	1.67	0
	K746	4	0
P63279	K153	1.95	0
P15873	K164	26.33	0
P41212	K99	7.67	0
P42858	K6	1.14	0
	K9	1.94	0
	K15	2.15	0
Q13485	K113	86	0
Q9BYV9	K202	3.82	1
	K421	3.28	1
Q9Y6K9	K309	2.36	0
O00429	K38	1.38	0
Q9UNR9	K560	10	1
Q05397	K152	3	0
P32457	K11	9.11	0
	K287	1.97	0
	K465	3	1
Q99497	K130	1.08	0
P16220	K285	2.58	1
	K304	1.83	1
P56817	K275	3.33	0
O15169	K857	2.47	1
	K860	1.7	0
Q9UER7	K630	1.81	0
	K631	1.7	0
Q15596	K788	6.67	1
P29590	K65	1.49	0
Q9H2X6	K1191	9	0
P03116	K514	2.67	0
Q9M0K4	K258	1.85	1
P03206	K12	1.5	0
P03209	K19	2.68	0
	K213	1.89	0

	K517	5	0
P61086	K13	2.67	0
P04591	K474	4.69	1
P00445	K69	1.48	0
Q07979	K322	2.05	1
P04456	K60	7	0
P11978	K1128	4.67	0
Q04322	K498	2.04	0
P06400	K720	6.67	0
O00180	K274	5	0
P54253	K16	4	1
	K194	3.19	1
	K610	3.21	0
	K697	4.67	0
	K746	3	0
P54132	K317	1.17	0
	K331	5	0
	K344	3.39	1
	K347	1.84	0
P42575	K77	8	0
Q16621	K368	3.23	1

Table S1 (B)

Table S2 - The validation performance of SUMOsp. For comparison, the *Sn*, *Sp*, *Ac*, and *CC* of the consensus motif Ψ-K-X-E are 74.48%, 98.16%, 97.21%, and 0.6689 respectively. (Refer to the manuscript for the definitions of *Sn*, *Sp*, *Ac*, and *CC*)

Table S2 (A) - The performance of self-consistency (Self)

Table S2 (B) - The performance of Leave-one-out validation (LOO)

Table S2 (C) - The performance of five-fold cross-validation (5-fold CV)

Self	Cut-Off	*Sn*	*Sp*	*Ac*	*CC*
GPS+MotifX	1.5	97.49	53.53	55.28	0.20
	4	89.12	80.07	80.43	0.32
	18	83.68	93.08	92.71	0.50
GPS	1.5	97.07	54.07	55.78	0.20
	4	75.31	83.87	83.53	0.29
	18	2.09	99.64	95.75	0.05
MotifX	1.5	82.85	93.38	92.96	0.50
	4	82.85	93.38	92.96	0.50
	18	82.85	93.38	92.96	0.50

Table S2 (A)

LOO	Cut-Off	*Sn*	*Sp*	*Ac*	*CC*
GPS+MotifX	1.5	97.74	53.47	55.24	0.20
	4	89.11	80.08	80.44	0.32
	18	83.68	93.08	92.71	0.50
GPS	1.5	97.32	54.01605	55.74	0.20
	4	75.22	83.88154	83.54	0.30
	18	2.09	99.63656	95.75	0.05
MotifX	1.5	82.85	93.38	92.96	0.50
	4	82.85	93.38	92.96	0.50
	18	82.85	93.38	92.96	0.50

Table S2 (B)

5-fold CV	Cut-Off	*Sn*	*Sp*	*Ac*	*CC*
GPS+MotifX	1.5	93.71	54.23	54.56	0.09
	4	87.40	80.61	80.66	0.15
	18	82.79	93.14	93.06	0.26
GPS	1.5	92.05	54.83	55.14	0.09
	4	69.83	84.47	84.34	0.13
	18	0	99.71	98.89	0.00
MotifX	1.5	82.79	93.38	93.29	0.27
	4	82.79	93.38	93.29	0.27
	18	82.79	93.38	93.29	0.27

Table S2 (C)

Figure S1 – The Receiver Operating Characteristic (ROC) curves of Self validation, Leave-one-out validation, and 5-fold cross validation. The areas under the ROC curves of 5-fold cross validation, self validation, and the leave-one-out validation are 0.7279, 0.7504, and 0.7513 respectively. (This is a partial ROC curve, due to the threshold-independent component, MotifX, incorporated in SUMOsp system.)