

Table S1: Cross Validation Results of Peptide-ligand ipHMMs

Domain	SMART Abr.	Sens ¹	Spec ²	Acc ³
Domain present in cyclins, TFIIB and Retinoblastoma	CYCLIN	0.00	0.70	0.68
Histone H3	H3	1.00	0.65	0.79
WD40 repeats	WD40	0.00	0.91	0.87
Alkaline phosphatase homologues	alkPPc	0.90	0.59	0.62
Actin	ACTIN	0.47	0.48	0.48
Insulin / insulin-like growth factor / relaxin family.	IIGF	0.77	0.62	0.65
Trypsin-like serine protease	TrypSPc	0.20	0.82	0.72
Beta-propeller repeat	PQQ	1.00	0.47	0.50
BPTI/Kunitz family of serine protease inhibitors.	KU	0.33	0.60	0.57
Src homology 2 domains	SH2	1.00	0.68	0.69
Kazal type serine protease inhibitors	KAZAL	0.46	0.80	0.72
Epidermal growth factor-like domain.	EGF	0.10	0.82	0.66
Alpha-lactalbumin / lysozyme C	LYZ1	0.80	0.66	0.67
Histone H2B	H2B	1.00	0.31	0.63
Caspase, interleukin-1 beta converting enzyme (ICE) homologues	CASc	0.00	0.55	0.37
Pancreatic ribonuclease	RNAsePc	0.82	0.64	0.67
EF-hand, calcium binding motif	EFh	0.57	0.86	0.79
Immunoglobulin	IG	0.00	1.00	0.96
Ligand binding domain of hormone receptors	HOLI	1.00	0.78	0.79
Serine/Threonine protein kinases, catalytic domain	STKc	0.41	0.91	0.88
Ricin-type beta-trefoil	RICIN	0.56	0.73	0.72
Extension to Ser/Thr-type protein kinases	STKX	0.00	0.60	0.59
Immunoglobulin C-Type	IGc1	0.45	0.65	0.62
Zinc-dependent metalloprotease	ZnMc	1.00	0.66	0.66
ATPases associated with a variety of cellular activities	AAA	0.00	1.00	0.97
Immunoglobulin V-Type	IGv	0.00	0.95	0.93
Ankyrin repeats	ANK	1.00	0.71	0.76
Serine Proteinase Inhibitors	SERPIN	0.29	0.73	0.72
Armadillo/beta-catenin-like repeats	ARM	0.20	0.92	0.83
Immunoglobulin C-2 Type	IGc2	0.56	0.73	0.71
Histone H4	H4	0.96	0.33	0.62
Histone 2A	H2A	0.88	0.43	0.58
Alpha-amylase domain	Aamy	0.19	0.77	0.75
Gelsolin homology domain	GEL	0.38	0.72	0.68
Rho (Ras homology) subfamily of Ras-like small GTPases	RHO	1.00	0.51	0.55
Src homology 3 domains	SH3	0.44	0.60	0.55
Summary		0.52	0.69	0.69

¹Sensitivity

²Specificity

³Accuracy = $\frac{TP+TN}{P+N}$