

Table S6: Validation Results of Nucleotide-ligand ipHMMs using generated Sequences

Domain	SMART Abr.	Sens <sup>1</sup>	Spec <sup>2</sup>	Acc <sup>3</sup>
Zinc finger	ZnF_C2H2	0.62	0.55	0.56
RNA recognition motif	RRM	0.76	0.66	0.68
Actin	ACTIN	0.88	0.73	0.76
Cyclic nucleotide-monophosphate binding domain	cNMP	0.63	0.68	0.67
DNA polymerase X family	POLXc	0.80	0.91	0.90
3'-5' exonuclease	35EXOc	0.68	0.73	0.72
Histidine kinase-like ATPases	HATPasec	0.46	0.68	0.63
Pancreatic ribonuclease	RNAsePc	0.80	0.62	0.66
Pumilio-like repeats	Pumilio	0.81	0.66	0.70
Serine/Threonine protein kinases, catalytic domain	STKc	0.37	0.78	0.72
DNA polymerase A domain	POLAc	0.79	0.77	0.77
ATPases associated with a variety of cellular activities	AAA	0.51	0.95	0.93
Homeodomain	HOX	0.79	0.60	0.66
Helix-hairpin-helix DNA-binding motif class 1	HhH1	0.85	0.80	0.80
Basic region leucine zipper	BRLZ	0.78	0.58	0.65
C4 zinc finger in nuclear hormone receptors	ZnF_C4	0.83	0.61	0.67
Rab subfamily of small GTPases	RAB	0.78	0.68	0.70
Helix-turn-helix lactose operon repressor	HTHLACI	0.88	0.62	0.70
Summary		0.72	0.70	0.72

<sup>1</sup>Sensitivity

<sup>2</sup>Specificity

<sup>3</sup>Accuracy =  $\frac{TP+TN}{P+N}$