Supplementary Table 1. Comparison between genes listed by CAPS GWS 2018 and other two GWS from 2017 and 2004 [4,5]. The differences observed between each GWS is mentioned in the 'Comments' column. To avoid ambiguity, same names of genes have been used in most cases instead of aliases.

CAPS 2018 GWS	2017 GWS	2004 GWS	Comments
		ACP1	Low Molecular Weight WPTP (acid phosphatase), not considered by CAPS GWS
	ALPI		Alkaline Phosphatase, not considered in CAPS GWS
	ALPL		Alkaline Phosphatase, not considered in CAPS GWS
	ALPP		Alkaline Phosphatase, not considered in CAPS GWS
	ALPPL2		Alkaline Phosphatase, not considered in CAPS GWS
CDC14A	CDC14A	CDC14A	
CDC14B	CDC14B	CDC14B	
	CDC25A	CDC25A	Belongs to different fold, not considered in CAPS GWS
	CDC25B	CDC25B	Belongs to different fold, not considered in CAPS GWS
	CDC25C	CDC25C	Belongs to different fold, not considered in CAPS GWS
CDKN3	CDKN3	CDKN3	
DNAJC6	DNAJC6		Not considered in 2004 GWS
DUPD1	DUPD1	DUPD1	
DUSP1	DUSP1	DUSP1	
DUSP10	DUSP10	DUSP10	
DUSP11	DUSP11	DUSP11	
DUSP12	DUSP12	DUSP12	
DUSP13	DUSP13	DUSP13Aa	Isoforms
		DUSP13Ba	Isoforms
DUSP14	DUSP14	DUSP14	
DUSP15	DUSP15	DUSP15	
DUSP16	DUSP16	DUSP16	
DUSP18	DUSP18	DUSP18	
DUSP19	DUSP19	DUSP19	
DUSP2	DUSP2	DUSP2	
DUSP21	DUSP21	DUSP21	
DUSP22	DUSP22	DUSP22	
DUSP23	DUSP23	DUSP23	
		DUSP24	Now referred to as DUSP26
		DUSP25	Now referred to as DUSP23
DUSP26	DUSP26	DUSP26	
DUSP27	DUSP27		Also referred to as DUPD1

DUSP28	DUSP28		Now referred to as DUSP26
DUSP3	DUSP3	DUSP3	
DUSP4	DUSP4	DUSP4	
DUSP5	DUSP5	DUSP5	
DUSP6	DUSP6	DUSP6	
DUSP7	DUSP7	DUSP7	
DUSP8	DUSP8	DUSP8	
DUSP9	DUSP9	DUSP9	
EPM2A	EPM2A	EPM2A	
	FIG4		Sac phosphatase, not considered in CAPS GWS
		EYA1	Asp-based phosphatase, not considered in CAPS GWS
		EYA1	Asp-based phosphatase, not considered in CAPS GWS
		EYA1	Asp-based phosphatase, not considered in CAPS GWS
		EYA1	Asp-based phosphatase, not considered in CAPS GWS
GAK	GAK		Not considered in 2004 GWS
	INPP5F		Sac Phosphatase
	LMWPTP		Low Molecular Weight PTP, not considered in CAPS GWS
MTM1	MTM1	MTM1	
MTMR1	MTMR1	MTMR1	
MTMR10	MTMR10	MTMR10	
MTMR11	MTMR11	MTMR11	
MTMR12	MTMR12	MTMR12	
MTMR13	MTMR13	MTMR13	
MTMR14	MTMR14	MTMR14	
		MTMR15	Not considered part of Myotubularin family, now referred to as FAN1
MTMR2	MTMR2	MTMR2	
MTMR3	MTMR3	MTMR3	
MTMR4	MTMR4	MTMR4	
MTMR5	MTMR5	MTMR5	
MTMR6	MTMR6	MTMR6	
MTMR7	MTMR7	MTMR7	
MTMR8	MTMR8	MTMR8	
MTMR9	MTMR9	MTMR9	
PALD1	PALD1		Not considered in 2004 GWS
PTEN	PTEN	PTEN	
PTP4A1	PTP4A1	PTP4A1	
PTP4A2	PTP4A2	PTP4A2	
PTP4A3	PTP4A3	PTP4A3	

PTPDC1	PTPDC1	PTP9Q22	Also referred to as PTPDC1
PTPMT1	PTPMT1		Discovered in 2005 [13], not considered in 2004 GWS
PTPN1	PTPN1	PTPN1	
PTPN11	PTPN11	PTPN11	
PTPN12	PTPN12	PTPN12	
PTPN13	PTPN13	PTPN13	
PTPN14	PTPN14	PTPN14	
PTPN18	PTPN18	PTPN18	
PTPN2	PTPN2	PTPN2	
PTPN20	PTPN20	PTPN20	
PTPN21	PTPN21	PTPN21	
PTPN22	PTPN22	PTPN22	
PTPN23	PTPN23	PTPN23	
PTPN3	PTPN3	PTPN3	
PTPN4	PTPN4	PTPN4	
PTPN5	PTPN5	PTPN5	
PTPN6	PTPN6	PTPN6	
PTPN7	PTPN7	PTPN7	
PTPN9	PTPN9	PTPN9	
PTPRA	PTPRA	PTPRA	
PTPRB	PTPRB	PTPRB	
PTPRC	PTPRC	PTPRC	
PTPRD	PTPRD	PTPRD	
PTPRE	PTPRE	PTPRE	
PTPRF	PTPRF	PTPRF	
PTPRG	PTPRG	PTPRG	
PTPRH	PTPRH	PTPRH	
PTPRJ	PTPRJ	PTPRJ	
PTPRK	PTPRK	PTPRK	
PTPRM	PTPRM	PTPRM	
PTPRN	PTPRN	PTPRN	
PTPRN2	PTPRN2	PTPRN2	
PTPRO	PTPRO	PTPRO	
PTPRQ	PTPRQ	PTPRQ	
PTPRR	PTPRR	PTPRR	
PTPRS	PTPRS	PTPRS	
PTPRT	PTPRT	PTPRT	
PTPRU	PTPRU	PTPRU	
		PTPRV	Now referred to as pseudophosphatase [14]

PTPRZ1	PTPRZ1	PTPRZ1	
RNGTT	RNGTT	RNGTT	
	SACM1L		Sac Phosphatase, not considered in CAPS GWS
SSH1	SSH1	SSH1	
SSH2	SSH2	SSH2	
SSH3	SSH3	SSH3	
	SSU72		Belongs to different fold, not considered in CAPS GWS
STYX	STYX	STYX	
STYXL1	STYXL1	STYXL1	
	SYNJ1		Sac Phosphatase, not considered in CAPS GWS
	SYNJ2		Sac Phosphatase, not considered in CAPS GWS
TNS1	TNS1	TNS	Now referred to as TNS1
TNS2	TNS2	TENC1	Now referred to as TNS2
TNS3	TNS3		
TPTE	TPTE	TPTE	
TPTE2	TPTE2	TPIP	Now referred to as TPTE2

Supplementary Table 2. Details of the thirty-seven domains the gene products containing tyrosine phosphatase domains associate with. Pfam, InterPro and/or CDD IDs have been provided for the domains in most cases. The entries in the 2nd column which are marked in bold correspond to domains which were not reported by the CAPS 2003 GWS. Asterisk symbols in the 1st column indicate the three domains newly found to associate with catalytic domains of tyrosine phosphatases.

S.No.	Domain Abbreviation	Name of the domain	Domain Database IDs	Function
1	**ALIX**	ALIX V-shaped domain or ALIX_LYPXL_bnd	Pfam: PF13949	Binding of the LYPxL motif of late HIV p6Gag and EIAV p9Gag proteins to this domain is necessary for viral budding
2	BRO1	BRO1-like domain	Pfam: PF03097	Known to have a role in endosomal targeting
3	**C1 or C1_1**	phorbol esters/diacylglycerol binding domain	Pfam: PF00130	Binds an important secondary messenger diacylglycerol (DAG), as well as the analogous phorbol esters. Phorbol esters can directly stimulate protein kinase C, PKC.

4	**C2**	C2 domain of PTEN tumour-suppressor protein	Pfam: PF10409	This domain associates across an extensive interface with the N-terminal phosphatase domain DSPc and possibly positions the catalytic part of the protein onto the membrane
5	Carbonic Anhydrase	Carbonic Anhydrase	Pfam: PF00194	Catalyse the reversible hydration of carbon dioxide to bicarbonate
6	CBM20	Carbohydrate-binding module	Pfam: PF00686	Known to have carbohydrate-binding activity
7	**CD45**	Leukocyte receptor CD45	Pfam: PF12567	It is required for T-cell activation through the antigen receptor
8	**CC**	Coiled-coil		A structural motif which aids in dimerization of myotubularins
9	CRAL_TRIO	Named after Cell RetinALdehyde-binding protein (CRALBP) and TRIO guanine exchange factor	Pfam: PF00650	It binds small lipophilic molecules
10	**DEK_C**	DEK C-terminal domain	Pfam: PF08766	A chromatin associated protein that is linked with cancers and autoimmune disease
11	**DENN**	DENN (AEX-3) domain or Differentially Expressed In Neoplastic vs Normal Cells	Pfam: PF02141	Occurs in several proteins involved in Rab- mediated processes or regulation of MAPK signalling pathways, function unclear
12	DNA-J	Chaperone DnaJ	Pfam: PF00226	DnaJ-domain is associated with hsp70 heat-shock system and is part of a chaperone (protein folding) system
13*	**Endomucin**	Endomucin	Pfam: PF07010	An early endothelial-specific antigen that is also expressed on putative hematopoietic progenitor cells
14	**FA**	FERM adjacent	Pfam: PF08736	This region is found adjacent to FERM domains and has

				been hypothesised to play a role in regulatory adaptation
15	FERM	FERM domain (F for 4.1 protein, E for ezrin, R for radixin and M for moesin)	Pfam: PF00373	Involved in localising proteins to the plasma membrane
16	fn3	Fibronectin type III domain	Pfam: PF00041	Involved in cell adhesion, cell morphology, thrombosis, cell migration, and embryonic differentiation
17	FYVE	FYVE zinc finger domain is named after the four cysteine-rich proteins: Fab 1 (yeast orthologue of PIKfyve), YOTB, Vac 1 (vesicle transport protein), and EEA1, in which it has been found	Pfam: PF01363	FYVE domains bind Phosphatidylinositol 3-phosphate and is implicated in vacuolar protein sorting and endosome function
18	Ig	Immunoglobulin domain	Pfam: PF00047	Involved in the recognition, binding, or adhesion processes of cells
19	Kinase	Protein kinase domain	Pfam: PF00069	Catalyse the transfer of the gamma phosphate from nucleotide triphosphates (often ATP) to one or more amino acid residues in a protein substrate side chain, resulting in a conformational change affecting protein function
20	**KIND**	Kinase non-catalytic C-lobe domain	Pfam: PF16474	Has evolved from a catalytic protein kinase fold and functions as an interaction domain
21	MAM	Meprin, A-5 protein, and receptor protein-tyrosine phosphatase mu	Pfam: PF00629	It occurs in several cell surface proteins and is likely to have an adhesive function. Known to play a role in homodimerization of protein-tyrosine phosphatase mu

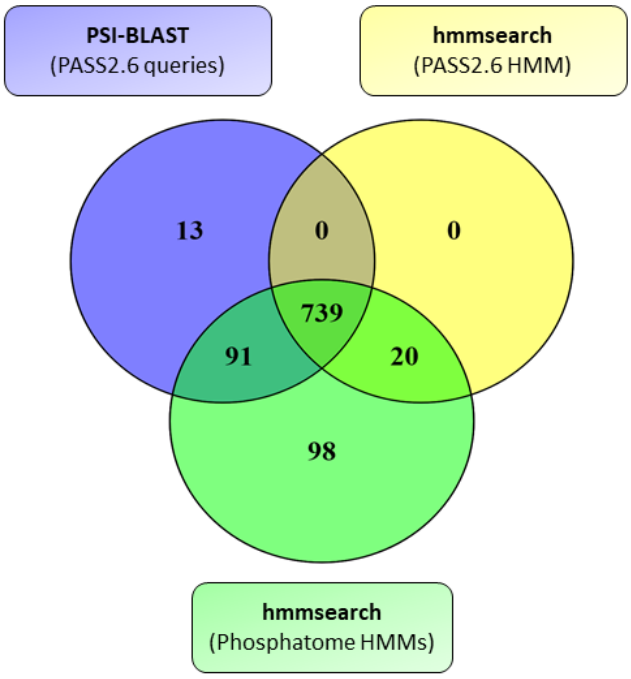
22	mRNA cap enzyme	mRNA capping enzyme	Pfam: PF01331	A mRNA guanyltransferase and RNA 5'-triphosphatase
23	PDZ	Name from the three proteins- post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1)	Pfam: PF00595	Play a key role in anchoring receptor proteins in the membrane to cytoskeletal components
24	**PDZ-binding motif**	PDZ-binding motif or PDZ domain binding		Predicted sequence motif which recognises PDZ domain containing proteins
25	**PH**	Pleckstrin Homology	Pfam: PF00169	Ooccurs in a wide range of proteins involved in intracellular signaling or as constituents of the cytoskeleton
26	GRAM	Glucosyltransferases, Myotubularins and other membrane-associated proteins	Pfam: PF02893	Implicated in intracellular protein-binding or lipid-binding signalling domain
27	PTB	Phosphotyrosine interaction domain (PTB/PID)	Pfam: PF00640	Function as adaptors or scaffolds to organise the signaling complexes involved in wide-ranging physiological processes including neural development, immunity, tissue homeostasis and cell growth
28*	**PTN13_u3**	Unstructured linker region on PTN13 protein between PDZ	Pfam: PF16599	Allows flexibility between the PDZ domains
29	**PTP_N**	Protein tyrosine phosphatase N terminal	Pfam: PF12453	Found in association with fn3 domains in proteins such as CD45 which limits growth signalling in haematopoietic cells
30	**Receptor-IA2**	Protein-tyrosine phosphatase receptor IA-2	Pfam: PF11548	Upon exocytosis, the cytoplasmic domain of IA-2 is cleaved and moves to the nucleus where it enhances transcription of the insulin gene

31	**RESP18**	Name derives from glucocorticoid-responsive protein regulated endocrine-specific protein 18 (RESP18)	Pfam: PF14948	N-terminal extracellular region of receptor-type tyrosine-protein phosphatases containing the protein-tyrosine phosphatase receptor IA-2 domain
32	**Rhodanese**	Rhodanese	Pfam: PF00581	A mitochondrial enzyme that detoxifies cyanide (CN^-) by converting it to thiocyanate (SCN^-)
33*	**Ricin or Ricin B-lectin**	Ricin-type beta-trefoil lectin domain	Pfam: PF00652	Domain with homology with Lectins, predicted carbohydrate binding function.
34	**Sec14**	Sec14p-like lipid-binding domain	CDD: cl15787	Found in secretory proteins, it is a lipid-binding domain
35	SH2	Src Homology 2) domain	Pfam: PF00017	SH2 domains allow proteins containing those domains to dock to phosphorylated tyrosine residues on other proteins
36	**SSH_N**	N-terminal domain conserved in slingshot (SSH) phosphatases	CDD: cd11652	May be involved in P-cofilin binding
37	TMH	Transmembrane Helix		

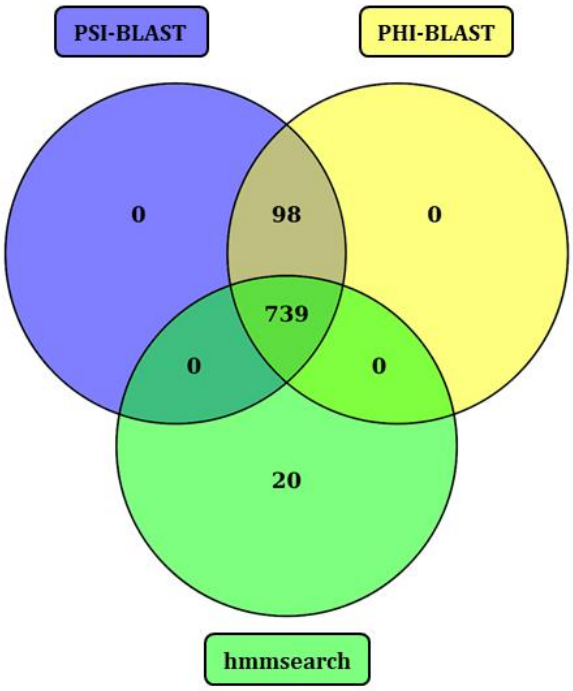
Supplementary Table 3: Few details about the three domains which are being reported for the first time to co-occur with tyrosine phosphatases. The E-values with which the domains were annotated by either of the domain annotation methods used, have been listed in columns 2 and 3 of the table. Under additional information, related literature evidence have been provided.

Domain Name	**CD-search**	**hmmsearch**	**Additional Information**
Ricin B-lectin	10E-4	10E-5	Within the Ricin B-like lectins superfamily (50370) in SCOPe, an existing structure of a mouse Ricin B-lectin domain (d1xhba1) was found. It belonged to the murine enzyme, UDP-GalNAc:polypeptide α-Nacetylgalactosaminyltransferase-T1 or ppGaNTase-T1. The motifs in its sequence were: Q-X-F, Q-X-W and Q-X-W.[28]
Endomucin	10E-5 (PTPRA) and 10E-10 (PTPRC)	-	Endomucin is known to be heavily glycosylated with O-linked glycans[30,31] and several serine and threonine residues were

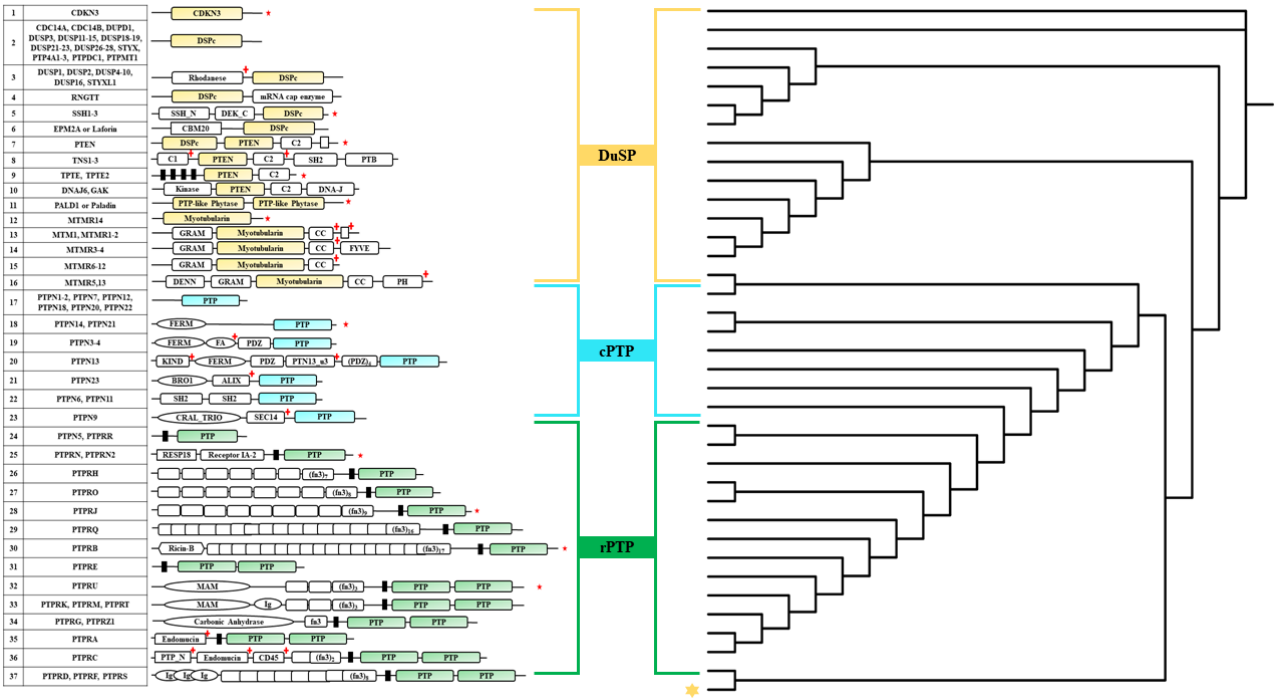
			also observed in the sequence stretches of PTPRA and PTPRC predicted to be Endomucin domain. PTPRA is known to be a regulator of the focal adhesion kinase[31] and it has been seen that the extracellular region of PTPRA, and not the catalytic domain, is required for modulation of cell surface expression of a neural adhesion molecule, NB-3.[32] Mutations in PTPRA are associated with conditions like schizophrenia, myocardial ischemia and others. CD45 or PTPRA is a hematopoietic tyrosine phosphatase and is critical for lymphocyte signalling. Role of PTPRC in adhesion in context of T cell adhesion has been well documented till date.[31,33,34]Mutations in this gene are implicated in severe autoimmune and infectious diseases.
PTN13_u3	10E-103	10E-73	A similar linker region, albeit shorter, has recently been reported in PTPN4 and this linker along with the adjacent PDZ domain were found to be crucial in allosteric regulation of the catalytic domain.[36] All the transcripts of the gene PTPN13 (ENSG00000163629) were annotated by the PTN13_u3 domain, in Ensembl.



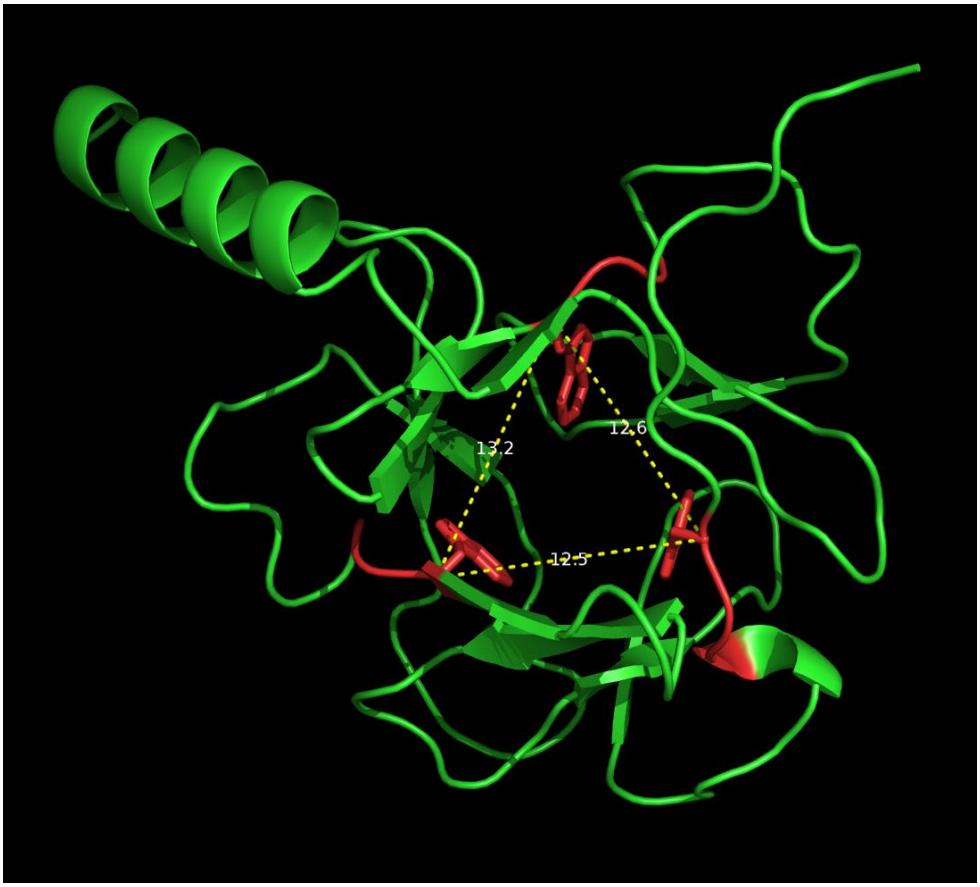
Supplementary Figure 1: Venn diagram representing the comparison of results obtained by carrying out sequence search using the PASS2.6 HMMs (hmmsearch) or query sequences (PSI-BLAST) with that obtained with HMMs obtained from the Phosphatome database (Chen et al., 2017). The 98 hits that were picked up by hmmsearch using Phosphatome HMM as query and not by other searches using PASS2.6 queries, correspond to inositol phosphatases and other lipid phosphatases, which do not come under the classification of PTPs that we have followed (Figure 1). We also compared hits obtained using different E-values and observed that maximum number of true positives are retained when E-value for PSI-BLAST and PHI-BLAST searches are 0.00001 and 0.1 for hmmsearch.



Supplementary Figure 2: Venn diagram representing hits picked up by multiple sequence search methods. A total of 739 hits were picked up by all three sequence search methods, 98 hits were picked up by the two profile-based search methods (PSI- and PHI-BLAST) and 20 by only the HMM based search method (hmmsearch).



Supplementary Figure 3: Tree constructed based on domain architecture dissimilarity, assessed using ADASS algorithm. The domain architectures 1-37 are listed to the left and the catalytic domains have been coloured according to the three main classes of tyrosine phosphatases- DuSP (yellow), cPTP (cyan) and rPTP (green). Each branch in the distance-based tree corresponds to domain architectures 1-37. The clusters of these three types of tyrosine phosphatases in the tree are correlated with the domain architectures listed to the left. The only DuSPs which cluster with other rPTPs (yellow star) are TPTE and TPTE2 and this could be due to the presence of transmembrane helices.



Supplementary Figure 4: The Ricin B-lectin domain of PTPRB: In the gene models of the PTPRB transcripts in Ensembl, we found that the predicted domain maps to the 2nd exon (ENSE00003551578) and the transcripts were annotated with the Ricin B-lectin domain as well. Homology model of the Ricin B-lectin domain of the PTPRB sequence was obtained using d1xhba1 structure as a template (Supplementary Table 3). Using MODELLER single template modelling and added restraints based on PSI-PRED results, the model was obtained. The PROSA Z-score was -3.48, falling in the region of published X-ray and NMR structures. The three tryptophans, which are part of the characteristic $(Q\text{-}X\text{-}W)_3$ motif, are almost equidistant from each other (C^{α} distance) and are marked in red.