

Supporting Information

SI Table 2. Filter criteria used for generating each FP dataset in Figure 3.

FP Rate (%)	XCorr> (2+, 3+, 4+)	dCn'	ppm filter
5	1.5, 1.5, 1.5	0.05	on
	1.5, 2.0, 2.5	0.18	off
2.5	1.5, 1.5, 1.5	0.09	on
	1.7, 2.4, 2.9	0.2	off
1	1.7, 1.8, 1.8	0.11	on
	1.9, 2.5, 2.9	0.25	off
0.5	1.8, 2.1, 2.9	0.3	on
	2.1, 2.5, 2.9	0.3	off

Supporting Information

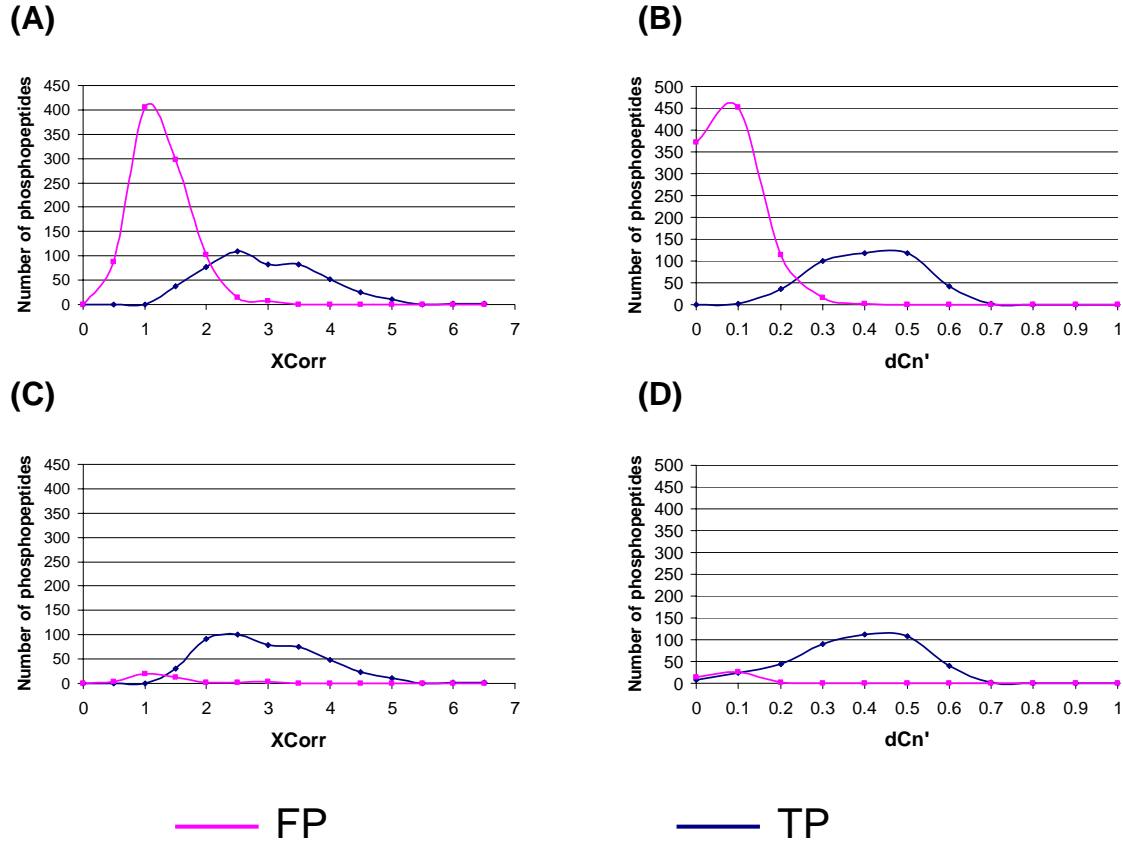
SI Table 3. Motifs extracted from our data set using the Motif-X algorithm²⁷.

Motif ^a	Corresponding kinase with known target sequence ^b	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
sP	Proline-directed	16.00	334	1208	36122	799758	6.12
sDXE	CK2	32.00	85	874	4250	763636	17.47
sDDD	CK2	39.31	27	789	459	759386	56.62
RRXs	PKA and PKC	32.00	54	762	2613	758927	20.58
sDXD	CK2	31.00	41	708	3008	756314	14.56
sEXE	CK2	32.00	48	667	4109	753306	13.19
RKXs	PKA and PKC	24.22	29	619	2673	749197	13.13
sXDD	CK2	25.76	29	590	3139	746524	11.69
RXXs	PKA and PKC	16.00	77	561	30586	743385	3.34
sXXD	CK2	12.61	68	484	37340	712799	2.68
sXDE	CK2	22.92	22	416	2988	675459	11.95
RXXSXXs	N/A	18.02	22	394	4289	672471	8.75
Es	CK2-like	8.78	52	372	37405	668182	2.50
Ds	CK2-like	7.40	45	320	36423	630777	2.44
RXs	PKA and PKC	6.22	31	275	24666	594354	2.72
sXXN	N/A	6.51	39	244	37654	569688	2.42
DXXs	CK1	6.14	31	205	30272	532034	2.66
tP	Proline-directed	16.00	114	237	27442	509522	8.93
RRXt	PKA and PKC	24.92	15	123	1194	482080	49.24
tXXE	CK2	13.72	34	108	32791	480886	4.62
RXXt	PKA and PKC	8.32	17	74	18007	448095	5.72

^aParameters used for motif extraction were: width (13), significance (10^{-6}), occurrence (23 for phosphorylated serine data set and 5 for phosphorylated threonine data set). “X” represents any amino acid residue. The phosphorylation sites are indicated by “s” or “t” in this table.

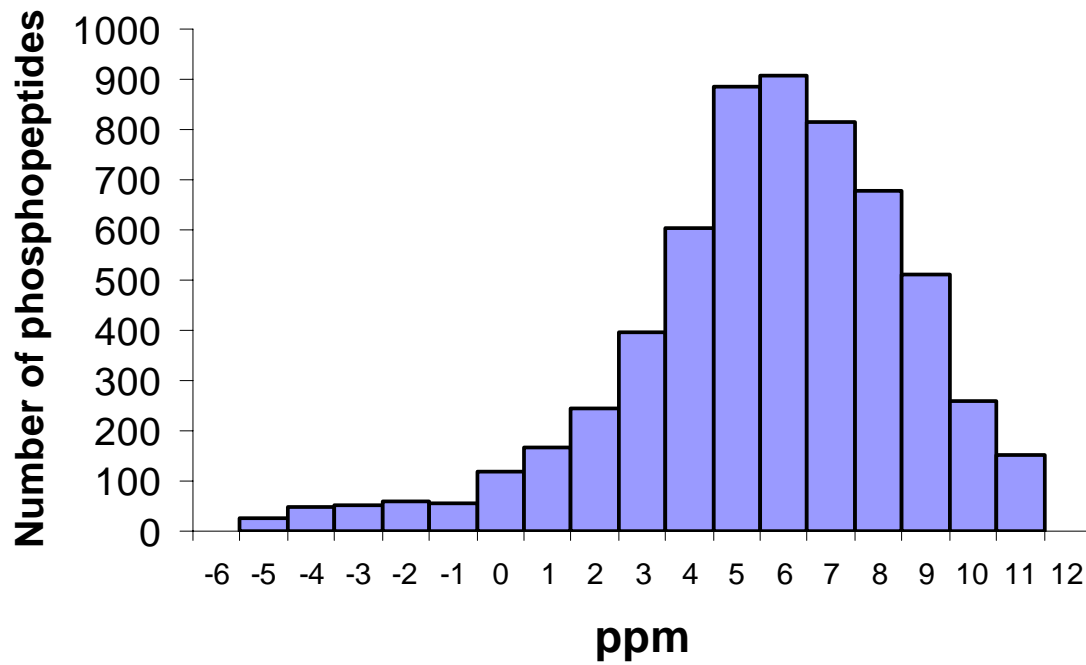
^bThe sequence specificity of kinases was obtained from: <http://www.hprd.org/>.

Supporting Information



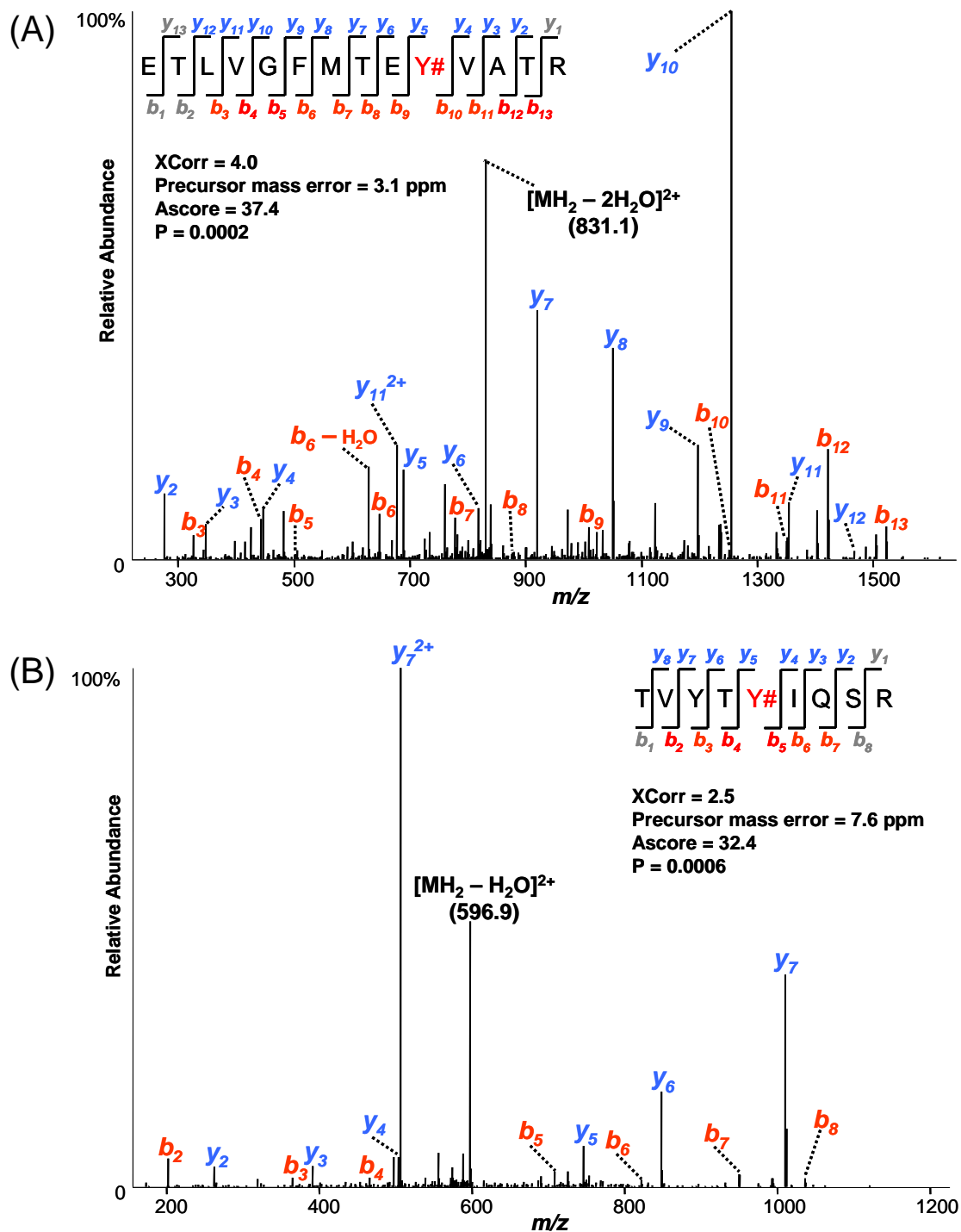
SI_Figure 1. XCorr and dCn' distributions of true- and false-positive phosphopeptide assignments with (C,D) and without (A,B) mass accuracy as a filter. MS/MS spectra from gel slice number 4 were searched with Sequest as described in the methods. For clarity, only doubly-charged phosphopeptide results are shown. TP and FP identifications were estimated based on the target/decoy database searching strategy within each bin. Note that, for example, accepting all phosphopeptides with XCorr > 2 in **Panel A** would include more than 400 TP results, but would also include more than 100 FP results. Using mass accuracy (**Panels C and D**) as a filter by requiring all matches to be within a 17 ppm window (5.5 ppm \pm 8.5) greatly affects the FP distributions, while the TP distributions remain almost entirely unaffected. The mass deviation distribution for all accepted phosphopeptides is shown in Figure SI_4.

Supporting Information



SI_Figure 2. Distribution of mass deviations of all identified phosphopeptides in Supplementary Table 1. These data centered at $\sim +5$ ppm and were not recalibrated. A window of 17 ppm was used for filtering based on this distribution.

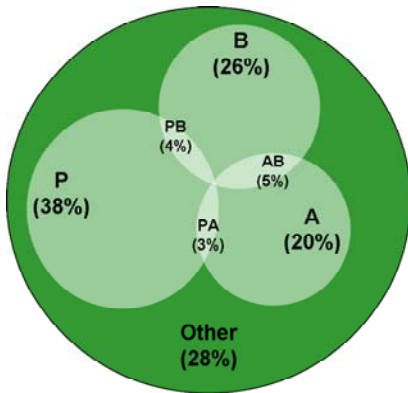
Supporting Information



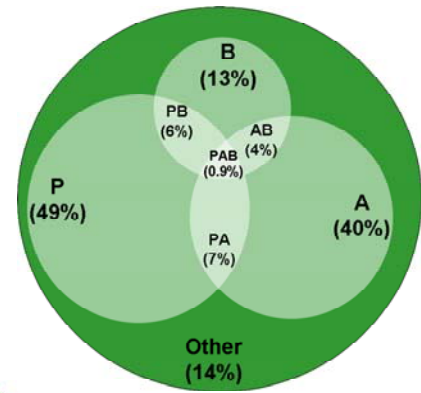
SI_Figure 3. Examples of two MS/MS spectra of tyrosine-phosphorylated peptides. There are no known tyrosine-specific kinases in yeast. However, there are dual-specificity kinases capable of tyrosine phosphorylation.

Supporting Information

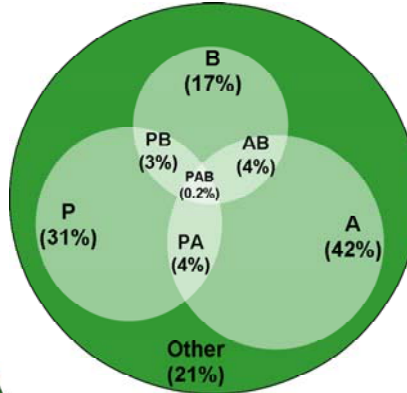
(A) Yeast (Gruhler *et al.*, 2005)



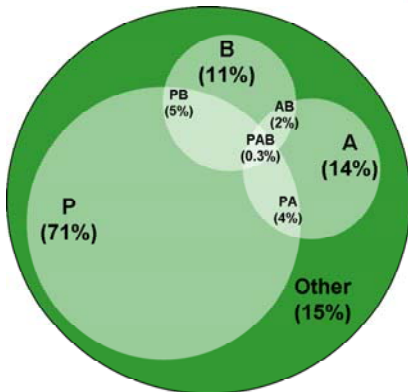
(B) Mouse brain



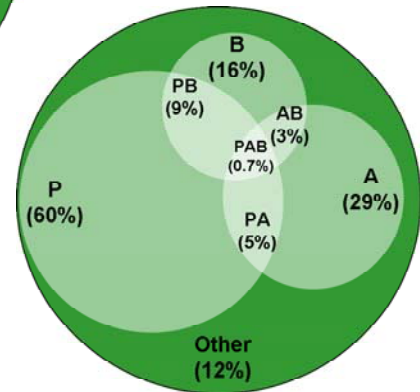
(C) This Data Set



(D) Noc-arrested human HeLa

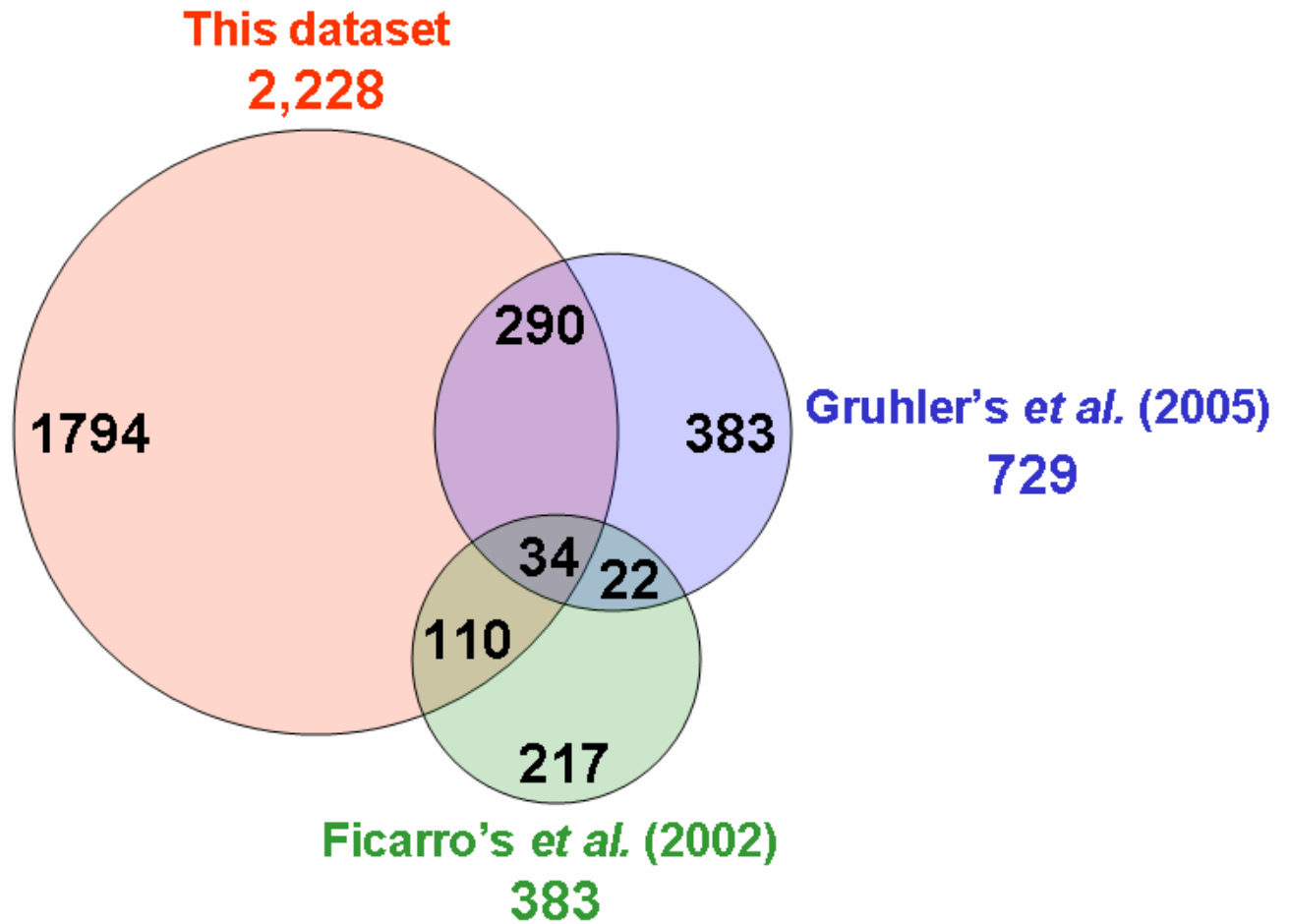


(E) Human HeLa nuclear fraction



SI Figure 4. Classification of phosphorylation events into 4 general sequence categories based on kinase specificities (P: proline-directed, A: acidiphilic, B: basophilic, O: others): **(A)** data set of α -factor arrested yeast¹¹, **(B)** data set of developing mouse brain⁹, **(C)** this data set, **(D)** data set of nocodazole-arrested human HeLa cells²⁵, **(E)** data set of human HeLa nuclear fraction¹⁰.

Supporting Information



SI Figure 5. Number of phosphorylation sites identified uniquely and in common between each IMAC-based yeast phosphoproteomic study.