

iGPS Manual

GPS algorithm with the interaction filter

Version 1.0.1 27/08/2012

Authors: Yu Xue & Jian Ren

Contact: Dr. Yu Xue, <u>xueyu@mail.hust.edu.cn</u>; Dr. Jian Ren, <u>renjian.sysu@gmail.com</u> The software is only free for academic research. The latest version of iGPS software is available from <u>http://igps.biocuckoo.org</u> Copyright (c) 2012. The CUCKOO Workgroup, DICP, HUST & SYSU. All Rights Reserved.

Index

Index	3
Statement	1
Introduction	2
Download & Installation	4
The Usage of iGPS	6
Direct Prediction	6
Batch Prediction	11
Visualization of Protein Phosphorylation Network with PNC 1.0	14
Search pre-calculated ssKSRs in EPNdb	17
References	21
Release Note	23

Statement

1. **Implementation**. The softwares of the CUCKOO Workgroup are implemented in JAVA (J2SE). Usually, both of online service and local stand-alone packages will be provided.

2. **Availability**. Our softwares are freely available for academic researches. For non-profit users, you can copy, distribute and use the softwares for your scientific studies. Our softwares are not free for commercial usage.

3. **GPS**. Previously, we used the GPS to denote our Group-based Phosphorylation Scoring algorithm. Currently, we are developing an integrated computational platform for post-translational modifications (PTMs) of proteins. We re-denote the GPS as Group-based Prediction Systems. This software is an indispensable part of GPS.

4. **Usage**. Our softwares are designed in an easy-to-use manner. Also, we invite you to read the manual before using the softwares.

5. **Updation**. Our softwares will be updated routinely based on users' suggestions and advices. Thus, your feedback is greatly important for our future updation. Please do not hesitate to contact with us if you have any concerns.

6. **Citation**. Usually, the latest published articles will be shown on the software websites. We wish you could cite the article if the software has been helpful for your work.

7. Acknowledgements. The work of CUCKOO Workgroup is supported by grants from Chinese 973 project (2010CB945400, 2012CB911200, 2012CB910101), and Chinese Natural Science Foundation (90919001, 30830036, 30900835, 31071154, 91019020, 31171263).

Introduction

Protein kinase (PK)-catalyzed phosphorylation is one of the most important and ubiquitous post-translational modifications (PTMs) of proteins. This process temporally and spatially modifies approximately 30% of all cellular proteins and plays a crucial role in regulating a variety of biological processes such as signal transduction and the cell cycle [1-6]. The human genome encodes 518 PK genes (approximately 2% of the genome), with different PKs exhibiting distinct recognition specificities; each PK modifies only a limited subset of substrates, thereby guaranteeing the fidelity of cell signaling [1-6]. It is accepted that short linear motifs (SLMs) around the phosphorylation sites (p-sites) provide primary specificity [2,6-9], and a variety of additional contextual factors, including co-localization, co-expression, co-complex, and physical interaction of the PKs with their targets, contribute additional specificity in vivo [10-15]. Aberrances of PKs or key substrates disrupt normal function, rewire signaling pathways, and are implicated in various diseases and cancers [4,16-19]. In this regard, the identification of kinase-specific p-sites and the systematic elucidation of site-specific kinase-substrate relations (ssKSRs) would provide a fundamental basis for understanding cell plasticity and dynamics and for dissecting the molecular mechanisms of various diseases, while the ultimate progress could suggest potential drug targets for future biomedical design [11,12,15].

Conventional experimental identification of ssKSRs, performed in a one-by-one manner, is labor-intensive, time-consuming and expensive. There are only 3,508 known kinase-specific p-sites in the 1,390 proteins collected in the Phospho.ELM 8.2 database (released in April 2009) [20]. In 2005, Ptacek *et al.* detected more than 4,000 *in vitro* kinase-substrate relations (KSRs) in *Saccharomyces cerevisiae* using protein chip technology, although the exact phosphorylation sites were not determined [3]. Recently, rapid advances in phosphoproteomics have provided an opportunity to systematically assess phosphorylation [1,21,22]. State-of-the-art high-throughput mass spectrometry (HTP-MS) techniques have the ability to detect thousands of p-sites in cells or tissues in a single experiment [1,21,23,24]. We have collected 145,646 eukaryotic p-sites, primarily from these large-scale assays (Supplemental Table S1); the regulatory PKs for 97.6% of these sites remain to be characterized.

Alternatively, the *in silico* prediction of ssKSRs can generate useful information for subsequent experimental manipulation. In 2001, Yaffe *et al.* developed the SLM-based software Scansite for the prediction of ssKSRs directly from protein primary sequences [10]. Later, the strategy was employed in a variety of kinase-specific predictors [25], including our group-based prediction system (GPS) program [26]. These tools may guarantee partially correct predictions for *in vitro* phosphorylation, but they are far from being

adequate for *in vivo* hits because the contributions of various contextual factors cannot be neglected. To address this problem, Linding *et al.* developed a predictor of NetworKIN by combining an SLM-based approach with network contextual information to predict *in vivo* ssKSRs, and a potential *in vivo* human phosphorylation network (HPN) was modeled by annotating the phosphoproteomic data [11,12].

In this work, we developed a software package of iGPS (GPS algorithm with the interaction filter, or in vivo GPS) mainly for the prediction of in vivo ssKSRs. Eukaryotic PKs were classified into a hierarchy with four levels: group, family, subfamily, and single PK [4]. Based on the hypothesis that similar PKs recognize similar SLMs, we selected a predictor in GPS 2.0 [26] for each PK and directly predicted the potential PKs for the non-annotated p-sites from the phosphoproteomic studies. Consequently, protein-protein interaction (PPI) information was used as the major contextual factor to filtrate potentially false-positive hits. The performance of iGPS was shown by critical evaluations and comparisons to be promising for the accurate prediction of in vivo ssKSRs. Based on the prediction results of iGPS, we modeled eukaryotic protein phosphorylation networks (PPNs) at different levels, including whole proteome. pathway and tissues/organs. By additionally computational analyses, we obtained a substantial number of potentially new observations, which can be subjected to further experimental manipulation. This study provides useful information for the understanding of the functional organization and diversity of eukaryotic phosphoproteomes at a systemic level and can be a model for analyzing other PTM-regulating proteomes.

ile Tools Help									
Phosphorylation		Second and the second	fic Kinase-substrate I						
🟱 🔲 🗂 Protein Kinase	Position	Code	Peptide	Matched ID	Gene Name	Kinase ID	Kinase Name	Interaction	
🕂 🔲 🗂 Serine/Threonine Kinase	>RApSpTIEM		1						
AGC	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P31749	AKT1	String	
	16	S	RSAIRRASTIEMPQQ	P26678	PLN	043930	PRKY	String	
- 🗹 🗋 AKT	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P22612	PKACg	String	
🗠 🗹 🚍 DMPK	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P17612	PKACa	Exp./String	
- 🗹 🗋 GRK	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P51817	PRKX	String	
	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P22694	PKACb	String	-
— 🗹 🗋 РКА	16	S	RSAIRRASTIEMPQQ	P26678	PLN PLN	D3DWF5	PKCb PKCa	String	-
— 🗹 🗋 РКВ	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P17252	PKCa	String String	-
- PKC	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P05129	PKCg	String	-
- 🗹 🗋 PKG	16	S	RSAIRRASTIEMPQQ RSAIRRASTIEMPOO	P26678	PLN	<u>Q05655</u>	PKCd	String	-
	16	S		P26678	PLN	P24723	PKCe	String	-
- 🗹 🗋 RSK	16	S	RSAIRRASTIEMPQQ	P26678	PLN	Q02156	PKCz	String	-
- 🗹 🗋 SGK	16	S	RSAIRRASTIEMPQQ RSAIRRASTIEMPQQ	P26678	PLN	Q05513	PKG2	String	-
← 🔲 🚍 CAMK	16	S	RSAIRRASTIEMPOO	P26678 P26678	PLN	Q13237 A5YM56	PKG1	String	-
	16	S	RSAIRRASTIEMPOO		PLN	075582	MSK1	String	H
— 🔲 🗋 СК1	16	S	RSAIRRASTIEMPOO	P26678 P26678	PLN	Q9UBS0	p70S6Kb	String	-
🗠 🔲 🔚 CMGC	16	S	RSAIRRASTIEMPQQ	P26678	PLN	075676	MSK2	String	-
🕶 🔲 🚍 STE		5	KOATKKADITEMPUU	P20078	I LIN	075070	more	Sung	•
			л						•
 Atypical Cther Tyrosine Kinase TK 		TDTPKAPVI APGpSPTEF SDNEEELV DAVLVDR PGTR VLpSETELK PSWSGSCR	ER						
	Options			<u></u>	c	onsole			
	Orga	nism H. sa	piens 🔻 F	Format PhosPe	р 🔻	PhosPe	p	Clear	
									_

The iGPS 1.0 is freely available at: <u>http://igps.biocuckoo.org</u>.

iGPS v1.0 User Interface

Download & Installation

The software of iGPS 1.0 has been implemented in JAVA, and can be installed on Windows, Mac OS X or Linux systems. The iGPS 1.0 is freely available for academic research at <u>http://igps.biocuckoo.org/down.php</u>. We recommend that users can download the latest release.

After downloading, please double-click on the install package to begin installation. Follow the user prompts through the installation. And snapshots of the setup program are shown below:



Select Start Menu Folder	
Where should Setup place the prog	ram's shortcuts?
Select the Start Menu folder in w program's shortcuts, then click N	hich you would like Setup to create th fext.
Group-based Prediction System	
7-Zip	
Accessories	
ACD Systems	
Administrative Tools	
Adobe Illustrator CS4 精简版	
Allway Sync	
CollabNet Subversion Server	
V Create shortcuts for all users	s
🔲 Don't create a Start Menu fold	ler
	<pre> Back Next > Canc</pre>





Click on the Finish button to complete the setup program.

The Usage of iGPS

Direct Prediction

The main propose of iGPS 1.0 was designed for the prediction of ssKSRs from the phosphoproteomic data. For convenience, the iGPS 1.0 allows users to input their data into the "TEXT form" for prediction. Three formats such as **PhosPep**, **ELM** or **FATSA** are adopted as below:

(1) PhosPep format: LVEDKPGpSR GEpSENAGTNQETR pSRpSNpSKSKPNLPSESR SKPNLPpSESR pTSEETISTVQEK pSLQPLAPR KDpSLLKPGLR pSGGQLPSLQEETTR

•••

Please note: The p must be annotated in front of the phosphorylation site. Only the annotated phosphorylation sites will be predicted. We **strongly recommend** users to input the data in this format, because it can be easily prepared from the phosphoproteomic experiments.

(2) ELM format:

acc sequence position code pmids kinases source entry_date Q99640

MLERPPALAMPMPTEGTPPPLSGTPIPVPAYFRHAEPGFSLKRPRGLSRSLPPPPPAKGSIPISRLFPPR TPGWHQLQPRRVSFRGEASETLQSPGYDPSRPESFFQQSFQRLSRLGHGSYGEVFKVRSKEDGRLYAVKR SMSPFRGPKDRARKLAEVGSHEKVGQHPCCVRLEQAWEEGGILYLQTELCGPSLQQHCEAWGASLPEAQ VWGYLRDTLLALAHLHSQGLVHLDVKPANIFLGPRGRCKLGDFGLLVELGTAGAGEVQEGDPRYMAPELLQ GSYGTAADVFSLGLTILEVACNMELPHGGEGWQQLRQGYLPPEFTAGLSSELRSVLVMMLEPDPKLRATAEA LLALPVLRQPRAWGVLWCMAAEALSRGWALWQALLALLCWLWHGLAHPASWLQPLGPPATPPGSPPCSLL LDSSLSSNWDDDSLGPSLSPEAVLARTVGSTSTPRSRCTPRDALDLSDINSEPPRGSFPSFEPRNLLSLFED TLDPT 426 S 12738781 PLK1 LTP 2004-12-31 00:00:00+01

• • •

Please note: This format was defined in the Phospho.ELM database [20]. All data should be delimited with **Tab**. Only phosphorylation sites with defined positions will be predicted.

(3) FASTA format:

. . .

Please note: All irregular words, including non-amino acid word (e.g., number) and blank, will be removed automatically. All of the potential phosphorylation sites in the sequences will be predicted. Here, we strongly recommend users **NOT** to input the data in this format, unless you can make sure the positions of real phosphorylation sites in the proteins. *Ab initio* prediction of ssKSRs merely from protein sequences with iGPS 1.0 will generate too much false positive hits.

Please choose a proper organism based on your input data. The iGPS can predict ssKSRs for five eukaryotic organisms, including *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*.

One example was selected for each format. Users can click on the "Example" button to access the examples.

	Predicte Position	d Site-s Code	Peptide		strate Relati Gene Name	ons Kinase ID	Kinase Name	Interaction	Predictor	Score	Cutoff
	Position	Code	Peptide	Matched ID	Gene Name	Kinase ID	Kinase Name	Interaction	Predictor	Score	Cutoff
 ↔ Atypical ↔ Other ♥ ☐ Tyrosine Kinase ↔ ☐ TK 											
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ANCLPSPT NGKPSQVI GLVAAYSG APGAIGPYp PSAPPpSP DLDEDELL SFPIKPVPp KVpSPVK EGMNPSVI Options	ESTDTPH IGAPGDS DpSDNEI DSQAVLVI PPPGTR GNLpSE SPSWSC DEYADpS	KAPVITLPS SPTEPAGQ EELVER DR DR TELK SSCR	YLER	.GER	PhosPep	Cons	ole Example		Clear	

ile Tools Help												
Phosphorylation	Predicted Site-specific Kinase-substrate Relations											
Protein Kinase ● ● ● <th>Position</th> <th>Code</th> <th>Peptide</th> <th>Matched ID</th> <th>Gene Name</th> <th>Kinase ID</th> <th>Kinase Name</th> <th>Interaction</th> <th>Predictor</th> <th>Score</th> <th>Cutot</th>	Position	Code	Peptide	Matched ID	Gene Name	Kinase ID	Kinase Name	Interaction	Predictor	Score	Cutot	
 C TKL Atypical C Other Tyrosine Kinase TK 	ANCLpSp1 NQKPSQV GLVAAYSQ APGAIGPY pSAPPpSF DLDEDELI SFPIKPVP KVpSPVK EGMNPSY	ESTDTPI NGAPGPS DpSDNE pSQAVLV PPPGTR LGNLpSE pSPSWS(DEYADpS	KAPVITLPS SPTEPAGQ EELVER DR TELK SSCR	YLER	GER	PhosPep	Cons	ole Example		Clear		

Choose one or more kinases from the Kinase Hierarchical Tree

Choose a **Threshold** that you need, while the default threshold is **Low**.

iGPS 1.0 - GPS algorithm with the int	eraction filte	er									
File Tools Help											
Phosphorylation	Predicte	d Site-s	specific ł	(inase-sub	strate Relati	ons					1
Protein Kinase Protein Kinase Image: Serine/Threonine Kinase Image: Comparison of the series Image: Comparison	Position	Code	Peptide	Matched ID	Gene Name	Kinase ID	Kinase Name	Interaction Interactinteraction Interaction Interaction Interaction I	Predictor	Score	Cutoff
	ANCLPSPT NQKPSQV GLVAAYSQ APGAIGPY pSAPPpSF DLDEDELL SFPIKPVP KVpSPVK EGMNPSY Options	ESTDTPH NGAPGpS DpSDNE pSQAVLVI PPPGTR _GNLpSE oSPSWS(DEYADpS	KAPVITLPS SPTEPAGQ EELVER DR ITELK SSCR SDEDQHDA	AYLER	GER Format	PhosPep Exp./String	Con	sole Example Network		Clear Submi	

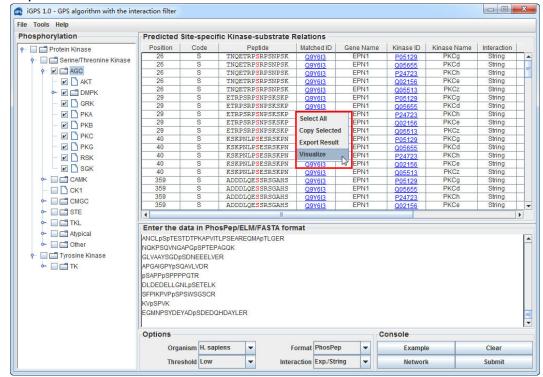
Also, the PPI information was adopted as a filter to reduce false positive hits. Both

experimentally identified and predicted PPI information was used, while the default parameter is all PPIs.

File Tools Help									
Phosphorylation	Predicted !	Site-spec	ific Kinase-substrate I	Relations					_
Protein Kinase	Position	Code	Peptide	Matched ID	Gene Name	Kinase ID	Kinase Name	Interaction	1
- Serine/Threonine Kinase	26	S	TNQETRPSRPSNPSK	Q9Y6I3	EPN1	P05129	PKCg	String	
	26	S	TNQETRPSRPSNPSK	Q9Y6I3	EPN1	Q05655	PKCd	String	
P ■ AGC	26	S	TNQETRPSRPSNPSK	Q9Y6I3	EPN1	P24723	PKCh	String	
- 🗹 🗋 AKT	26	S	TNQETRPSRPSNPSK	Q9Y6I3	EPN1	Q02156	PKCe	String	
	26	S	TNQETRPSRPSNPSK	Q9Y6I3	EPN1	Q05513	PKCz	String	
	29	S	ETRPSRPSNPSKSKP	Q9Y6I3	EPN1	P05129	PKCg	String	
- 🗹 🗋 GRK	29	S	ETRPSRPSNPSKSKP	Q9Y6I3	EPN1	Q05655	PKCd	String	
— 🗹 🗋 PKA	29	S	ETRPSRPSNPSKSKP	Q9Y6I3	EPN1	P24723	PKCh	String	
— 🔽 🗋 РКВ	29	S	ETRPSRPSNPSKSKP	Q9Y6I3	EPN1	Q02156	PKCe	String	
PKC	29	S	ETRPSRPSNPSKSKP	Q9Y6I3	EPN1	Q05513	PKCz	String	
	40	S	KSKPNLPSESRSKPN	Q9Y6I3	EPN1	P05129	PKCg	String	
- 🗹 🗋 PKG	40	S	KSKPNLPSESRSKPN	Q9Y6I3	EPN1	Q05655	PKCd	String	
- 🗹 🗋 RSK	40	S	KSKPNLPSESRSKPN	Q9Y6I3	EPN1	P24723	PKCh	String	
SGK	40	S	KSKPNLPSESRSKPN	Q9Y6I3	EPN1	Q02156	PKCe	String	
	40	S	KSKPNLPSESRSKPN	Q9Y6I3	EPN1	Q05513	PKCz	String	
🗢 🔲 🚍 CAMK	359	S	ADDDLQESSRSGAHS	Q9Y6I3	EPN1	P05129	PKCg	String	
— 🔲 🗋 СК1	359	S	ADDDLQESSRSGAHS	Q9Y6I3	EPN1	Q05655	PKCd	String	
	359	S	ADDDLQESSRSGAHS	Q9Y6I3	EPN1	P24723	PKCh	String	
	359	S	ADDDLQESSRSGAHS	Q9Y6I3	EPN1	Q02156	PKCe	String	
← □	4		JII.						₽
 Atypical C ther Tyrosine Kinase K 		TDTPKAPVI APGpSPTEF SDNEEELV DAVLVDR PGTR ILpSETELK 'SWSGSCR	ER						
	Options					Console			
	Orga	nism H. sa	piens 🔻 F	ormat PhosPe	p 🔻	Exampl	e	Clear	
		nism H. sa shold Low		ormat PhosPe action Exp./Str		Exampl	-	Clear	_

Click on the **Submit** button, then the predicted ssKSRs will be shown.

Then please click on the **RIGHT** mouse button in the results form.



You can use the "Select All" and "Copy Selected" to copy the selected results into Clipboard. Then please copy the results into a file, e.g., an EXCEL file for

further consideration. Also, you can choose "**Export Result**" to export the results into a tab-delimited text file.

If you choose the **Visualize** function, the given protein, its phosphorylation sites and phospho-peptides will be visualized with DOG (<u>Domain Graph</u>, Version 2.0), an illustrator of protein domain structures.



Batch Prediction

We also provide an alternative approach for processing multiple files. If each file is large, the **Batch Predictor** will be convenient for users.

The inputted data should be prepared in **PhosPep**, **ELM** or **FATSA** format. The mixed formats are not permitted. To run the Batch Predictor just select the **Batch Predictor** option in the **Tools** menu.

PKs	Sequence File List	Result File List
 Protein Kinase Serine/Threonine Kinase AGC CAMK CAMK CK1 CK1 CK1 TKL Atypical Tyrosine Kinase TK 	>>	
	Remove All Remove Add File Options	Export Folder >> Console
	Organism H. sapiens - Format FASTA	
	Threshold Low Interaction Exp./St	tring 👻 Submit

Click on the Add File button and add one or more files in your hard disk.

PKs	Sequence File List	Result File List
Protein Kinase AGC AGC CAMK CK1 CK1 CMGC STE TKL TKL TKL TK TK	查看: PhosPep PhosPep1.txt PhosPep2.txt PhosPep3.txt PhosPep4.txt PhosPep5.txt	▼ ⑤ ① 288 E ● 2 bt" "PhosPep3.bt" "PhosPep4.bt" "PhosPep5.bt" ● 11开 取消
	Remove All Remove Add File Options	Export Folder >> Console

Ks	Sequence File List	nî î	Result File List	
 Protein Kinase Serine/Threonine Kinase AGC AGC CAMK CK1 CK1 CMGC TKL TKL Atypical Other Tyrosine Kinase TK 	E:IPhosPepIPhosPep1.txt E:IPhosPepIPhosPep3.txt E:IPhosPepIPhosPep3.txt E:IPhosPepIPhosPep4.txt E:IPhosPepIPhosPep5.txt	>>		
	Remove All Remove Add File Options		Export Folder E:\PhosPep	>>
	Organism H. sapiens 🔻 Format	FASTA	A	Clear

The name of added files will be shown in the Sequence File List

The output directory of prediction results should also be defined. Please click on the ">>" button to specify the export file folder.

E:\PhosPep\PhosPep1.txt E:\PhosPep\PhosPep2.txt E:\PhosPep\PhosPep3.txt E:\PhosPep\PhosPep4.txt
E:IPhosPepIPhosPep5.txt >>
Remove All Remove Add File Export Folder E\PhosPep >> Options Console
Organism H. sapiens 👻 Format PhosPep 👻 Clear

Choose one or more kinases from **Kinase Hierarchical Tree**, and then pick a **Threshold**, click on the **Submit** button, then the **Batch Predictor** will begin to process all of the sequence files that have been added to the list. The results of predictions will be exported to the **Result Export Folder**, and the name of result files will be shown in the **Result File List**.

PKs	Sequence File List Result	File List
Protein Kinase Serine/Threonine Kinase AGC AGC	E:IPhosPepiPhosPep2.txt E:IPhosP E:IPhosPepiPhosPep3.txt E:IPhosP E:IPhosPepiPhosPep4.txt E:IPhosP	PepiPhosPep1.iGPS PepiPhosPep2.iGPS PepiPhosPep4.iGPS PepiPhosPep4.iGPS PepiPhosPep5.iGPS
🍐 🗌 🚍 ТК	Remove All Remove Add File Export 1 Options	Folder EAPhosPep >> Console
	Organism H. sapiens 💌 Format PhosPep	✓ Clear

Visualization of Protein Phosphorylation Network with PNC 1.0

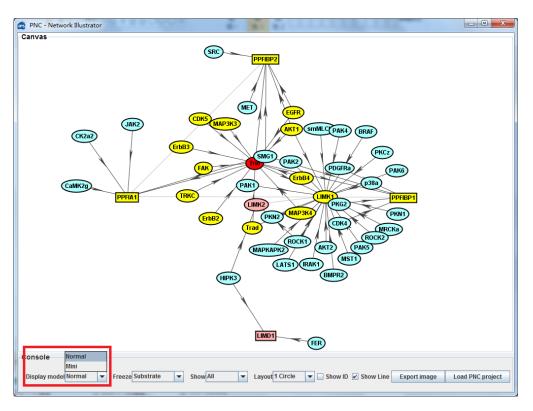
After prediction, you can click on the **Network** button or use the right button menu option to build a protein phosphorylation network (PPN).

hosphorylation	Predicted	Site-sneci	fic Kinase-substrate	Pelations					
- C C Protein Kinase	Position	Code	Peptide	Matched ID	Gene Name	Kinase ID	Kinase Name	Interaction	-
			replice	Matericano	oche Marrie	Tundoe ID	Tundoc Hame	meracaon	-
👇 🔲 🗂 Serine/Threonine Kinase	16	S	RSAIRRASTIEMPOO	P26678	PLN	P31749	AKT1	String	-
Y → AGC	16	S	RSAIRRASTIEMPOO	P26678	PLN	043930	PRKY	String	-
	16	S	RSAIRRASTIEMPOO	P26678	PLN	P22612	PKACq	String	-
	16	S	RSAIRRASTIEMPOO	P26678	PLN	P17612	PKACa	Exp./String	-
← 🗹 🗂 DMPK	16	S	RSAIRRASTIEMPOO	P26678	PLN	P51817	PRKX	String	-
- 🗹 🗋 GRK	16	S	RSAIRRASTIEMPOO	P26678	PLN	P22694	PKACb	String	
- PKA	16	S	RSAIRRASTIEMPOO	P26678	PLN	D3DWF5	PKCb	String	-
- 🔽 🗋 РКВ	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P17252	PKCa	String	
	16	S	RSAIRRASTIEMPOO	P26678	PLN	P05129	PKCg	String	-
— 🗹 🗋 РКС	16	S	RSAIRRASTIEMPQQ	P26678	PLN	Q05655	PKCd	String	
- 🗹 🎦 PKG	16	S	RSAIRRASTIEMPQQ	P26678	PLN	P24723	PKCh	String	
	16	S	RSAIRRASTIEMPQQ	P26678	PLN	Q02156	PKCe	String	
	16	S	RSAIRRASTIEMPQQ	P26678	PLN	Q05513	PKCz	String	
- 🗹 🗋 SGK	16	S	RSAIRRASTIEMPOO	P26678	PLN	Q13237	PKG2	String	
🔶 🔲 🚍 CAMK	16	S	RSAIRRASTIEMPOO	P26678	PLN	A5YM56	PKG1	String	
- 🔲 🗋 СК1	16	S	RSAIRRASTIEMPQQ	P26678	PLN	075582	MSK1	String	
	16	S	RSAIRRASTIEMPQQ	P26678	PLN	Q9UBS0	p70S6Kb	String	
	16	S	RSAIRRASTIEMPQQ	P26678	PLN	075676	MSK2	String	
🗠 🔲 🚍 STE	•		, II				, 		•
👇 🔲 🚍 TKL	Enter the	data in Ph	osPep/ELM/FASTA for	mat					_
 ← ☐ Atypical ← ☐ Other ← ☐ Tyrosine Kinase ← ☐ TK 	LVEDKPGpSI GEpSENAGTI pSRpSNpSK3 SKPNLPpSE3 pTSEETISTV0 pSLOPLAPR KDpSLLKPGI pSGGQLPSLI TADAPSEPAA	NQETR SKPNLPSES SR QEK LR QEETTR QSPHQR	R						
	KDpSEEEVSI								
	Options				C	onsole			
	Options	anism H. sa	piens 🔻 I	Format PhosPe		onsole PhosPe	p	Clear	

Diagram:

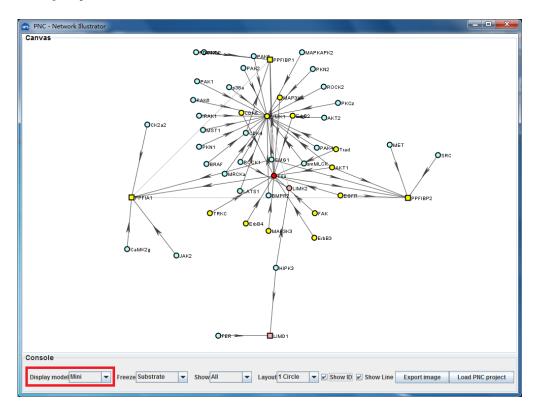
	Normal display model	Mini display model
Substrate in Inpu	ıt 📃	
Kinase in Input	\bigcirc	0
Regulatory Kinas	se 🔵	0
Selected Unit		•
Neighbor Unit	\bigcirc	•

The orientation was defined as Kinase -> Substrate



Normal display model:

Mini display model:

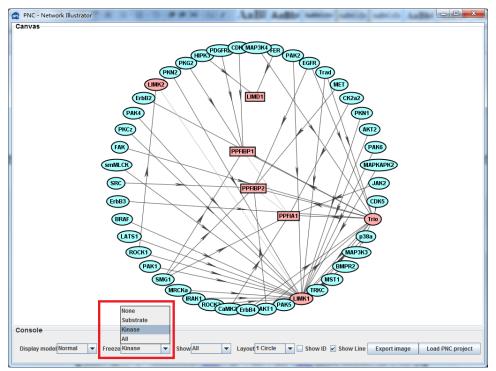


Freeze menu options

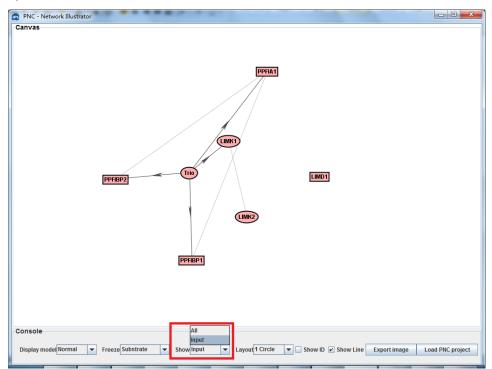
Kinase: freeze input kinase and related kinase. If you choose a **Circle Layout** model, the kinases will be located as hollow circles.

Substrate: freeze the input substrate. If you choose a **Circle Layout** model, the substrates will be located as hollow circles.

All: freeze all display units. None: release all display units.



If you choose the **Input** option in the **Show** menu, only the kinases and substrates in the input data will be shown.



Search pre-calculated ssKSRs in EPNdb

From public databases and the scientific literature, we collected 145,646 experimentally identified phosphorylation sites in 28,457 substrates, with 14,534, 5,555, 15,622, 49,119 and 60,816 phosphorylation sites in *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *M. musculus* and *H. sapiens*, respectively. With the PPI information, we predicted 186,922 (total PPIs) and 34,873 (experimental PPIs) ssKSRs. With these prediction results, we constructed the EPNdb 1.0. The users can search this database for pre-calculated ssKSRs.

Tools												
nase-su	bstrate relatio	nship	Interaction info	rmation								
ID	Position	Code	Peptide	M. ID	M. Position	Gene Name	Kinase Na	Kinase ID	Interaction	Predictor	Score	Cuto
										-		-
												-
				-	-	2						
												-
							-	-				-
						2	<u>, </u>		2	<u>, </u>		-
												-
										-		
												-
									-			-
				2		2						-
	1 1				1				ka sa			
ter the	e sequences	s in FAS	STA/PhosPep/	ELM form	at							
FHC3												
TQFNK	GPSYGLSAEV	KNRLLS	KYDPQKEAELRT	WIEGLTGL	SIGPDFQKGLI	KDGTILCTLM	NKLQLGSVP	KINRSMQNW	HQLENLSNF	IKAMVSYGMI	NPVDLFEAN	DLFESC
FHE4												
STQFNK	GPSYGLSAEV	KNRLLS	KYDPQKEAELRT	WIEGLTGL	SIGPDFQKGLI	KDGTILCTLM	NKLQPGSVP	KINRSMQNV	VHQLENLSN	IKAMVSYGM	NPVDLFEAN	DLFES
5962												
MDVLP	ILKEKVAYLSG	GRDKRO	GPILTFPARSNH	DRIRQEDL	RRLISYLACIP	SEEVCKRGF	TVIVDMRGS	WDSIKPLLK	ALQESFPCCI	HVALIIKPDN	FWQKQRTN	FGSSKF
nsole												
	l. sapiens						matFASTA					
aniem H	eanione	-	Threshold Low	The second secon	teraction All	Ter For	matEASTA	T FAS	CL CL	ear Su	ubmit	Networ

Please input data into the "TEXT form".

The data should be prepared in FATSA, PhosPep or ELM format.

Please choose a proper organism based on your input data.

Tools			ukaryotic Phosp									
	bstrate relatio	nship	Interaction info	rmation								
ID	Position	Code	Peptide	M. ID	M. Position	Gene Name	Kinase Na	Kinase ID	Interaction	Predictor	Score	Cutoff
				2								
				2		-						
				2	(y						· · · · · · · · · · · · · · · · · · ·	
				2	<u></u>						-	
				2								1
				2	19 1				,		-	
				2								
				2								
				2								1
				2								
				2	<i></i>						-	
SFHC3 STQFNK SFHE4 STQFNK 25962 H MDV M	GPSYGLSAEV GPSYGLSAEV sapiens . musculus . melanogaste	KNRLLS	STA/PhosPep/ SYDPQKEAELRT SYDPQKEAELRT SGPILTFPARSNH	WIEGLTGL	SIGPDFQKGLI	KDGTILCTLM	NKLQLGSVP) NKLQPGSVPI	KINRSMQNV	HQLENLSNF	IKAMVSYGM	NPVDLFEAN	NDLFESG
S.	. elegans . cerevisiae											
	. sapiens	-	Threshold Low	T Int	eraction All	- For	mat FASTA	▼ FAS		ar Su	ıbmit	Network

Choose a **Threshold** what you need, while the default is **Low**.

👩 EPNdb 1.	0 - The Data	base of Euka	aryotic Phosp	horylaton N	etwork							
File Tools												
Kinase-substrate relationship Interaction information												
ID	Position	Code	Peptide	M. ID	M. Position	Gene Name	Kinase Na	Kinase ID	Interaction	Predictor	Score	Cutoff
	2	p		p	-		<u>, </u>	e			-	
							,					
				2 2		,	2					
				2							-	
				2								
							,				-	
	2	p	,, ,	p		,	,	e				
		2		2			-					
Enter the	sequence	s in FASTA	/PhosPep/	ELM forma	at .							I
	VENDIVITY	I /II OWNER	wowooder.	North Ho						Concinee.		
>Q6FHC3 MSSTQFNK0	SPSYGL SAEV	KNRLLSKY	POKEAELRT	WIEGLTGLS	IGPDFQKGL		NKLQLGSVP	KINRSMONN	HOLENLSNE	IKAMVSYGM		DLEESGN
>Q6FHE4												
	GPSYGLSAEV	KNRLLSKYE	PQKEAELRT	WIEGLTGLS	IGPDFQKGL	COGTILCTLM	NKLQPGSVP	KINRSMQNV	HQLENLSN	TKAMVSYGM	NPVDLFEAN	DLFESGN
>075962 MKAMDVLPIL	KEKVAVI SG	CROKROCE				SEEVOKROE	TVIVDMRGSK	WDSIKPLLK				CSSKEEF -
	ENERGYATEOO	I CITE I TO CITU			(TEIOTE/IOII	OLEVOITION	TTTDMICOOI	CHEORY EER	iedeoi i ooi	TWALING DIV	The Cartoart Critic	v v
•			High									
Console			Mediu	Im								
Organism H.	sapiens	Th	reshold Low	▼ Inte	eraction All	▼ For	mat FASTA	▼ FAS	TA CI	ear Su	ıbmit	Network
-									- 70			

The protein-protein interaction (PPI) information was used as the major filter to reduce false positive predictions.

Interaction menu options

Exp.: filter the predictions with the experimentally identified physical interactions between PKs and substrates.

All: filter the predictions with both STRING and experimental PPI.

EPNdb 1.0 - The Database of	Eukaryotic Phosphorylaton Net	work		
File Tools				
Kinase-substrate relationship	Interaction information			
Protein_A ID	Protein_A Name	Protein_B ID	Protein_B Name	Interaction
-				
-				
-				
Enter the sequences in FA	ASTA/PhosPep/ELM format			
Console		Exp./String		
Organism H. sapiens 💌	Threshold Low 💌 Interact	Exp. tion Exp./String - Format FAST	A FASTA	Clear Submit Network
organisti n. sapiens	Interaci	round rai	FASTA	Sublin Network

Click on the **Submit** button, then the predicted site-specific kinase-substrate relations will be shown.

(inase-su	ibstrate rela	tionship	Interaction informatio	n								
ID	Position	Code	Peptide	M. ID	M. Position	Gene Na	Kinase N	Kinase ID	Interaction	Predictor	Score	Cutoff
9UGP4	272	S	TASSORVSPGLPSPN	Q9UGP4	272	LIMD1	HIPK3	Q9H422	Exp./String	CMGC/DYRK	5.8	1.17
9UGP4	277	S	RVSPGLPSPNLENGA	Q9UGP4	277	LIMD1	HIPK3	Q9H422	Exp./String	CMGC/DYRK	1.6	1.17
9UGP4	384	S	LGTGPKLSPTSLVHP	Q9UGP4	384	LIMD1	HIPK3	Q9H422	Exp./String	CMGC/DYRK	1.3	1.17
9UGP4	294	Т	VGPVQPRTPSVSAPL	Q9UGP4	294	LIMD1	HIPK3	Q9H422	Exp./String		2.6	1.17
9UGP4	296	S	PVQPRTPSVSAPLAL	Q9UGP4	296	LIMD1	HIPK3	Q9H422	Exp./String	CMGC/DYRK	1.7	1.17
9UGP4	424	S	VLLDSPSSPRVRLPC	Q9UGP4	424	LIMD1	HIPK3	Q9H422	Exp./String	CMGC/DYRK	3.0	1.17
9UGP4	272	S	TASSQRVSPGLPSPN	Q9UGP4	272	LIMD1	FER	P16591	String	TK/Fer/Fer	3.25	3.01
9UGP4	527	Y	FCEEDFLYSGFQQSA	Q9UGP4	527	LIMD1	FER	P16591	String	TK/Fer/Fer	5.25	3.01
953667	508	Т	PDRKKRYTVVGNPYW	P53667	508	LIMK1	PAK5	Q9P286	String	STE/STE20	6.333	1.22
53667	274	S	GPETSPLSSPAYTPS	P53667	274	LIMK1	PAK5	Q9P286	String	STE/STE20	1.233	1.22
53667	296	S	RQKPVLRSCSIDRSP	P53667	296	LIMK1	PAK5	Q9P286	String	STE/STE20	1.8	1.22
53667	508	Т	PDRKKRYTVVGNPYW	P53667	508	LIMK1	PAK6	Q9NQU5	String	STE/STE20	6.333	1.22
53667	274	S	GPETSPLSSPAYTPS	P53667	274	LIMK1	PAK6	Q9NQU5	String	STE/STE20	1.233	1.22
53667	296	S	RQKPVLRSCSIDRSP	P53667	296	LIMK1	PAK6	Q9NQU5	String	STE/STE20	1.8	1.22
53667	508	Т	PDRKKRYTVVGNPYW	P53667	508	LIMK1	MRCKa	Q5VT25	String	AGC/DMPK	6.121	1.21
53667	310	S	PGAGSLGSPASQRKD	P53667	310	LIMK1	p38a	Q16539	Exp./String		2.218	0.83
53667	210	S	GVDPGCMSPDVKNSI	P53667	210	LIMK1	p38a	Q16539	Exp./String		1.351	0.83
253667	302	S	RSCSIDRSPGAGSLG	P53667	302	LIMK1	p38a	Q16539	Exp./String		1.862	0.83
253667	229	Т	RILEINGTPIRNVPL	P53667	229	LIMK1	p38a	Q16539	Exp./String	CMGC/MAPK	1.004	0.83
953667	313	S	GSLGSPASORKDLGR	P53667	313	LIMK1	p38a	Q16539	Exp./String		0.874	0.83
253667	508	Т	PDRKKRYTVVGNPYW	P53667	508	LIMK1	PKN2	Q16513	String	AGC	2.126	1.59
953667	298	S	KPVLRSCSIDRSPGA	P53667	298	LIMK1	PKN2	Q16513	String	AGC	2.147	1.59
nter the	e sequen	ces in FA	STA/PhosPep/ELM f	ormat								
9BT23			KDGLFRVDKGAGNNPEF									
953667			VGRSCGQRIYDGQYLQA									
53671					THINNELLE		TIERDGQL	FOLIOOOL	KIGESCHG	CSEQUINCEVM	VAGELNIN	PEOFICEI
												•
onsole												

The interactions among input substrates also displayed.

inase-substrate relationship	Interaction information			
Protein_A ID	Protein_A Name	Protein_B ID	Protein_B Name	Interaction
<u>P53667</u>	LIMK1	<u>P53671</u>	LIMK2	Exp.
013136 PPFIA1		<u>Q86W92</u>	PPFIBP1	Exp./String
<u>Q13136</u>	PPFIA1	<u>Q8ND30</u>	PPFIBP2	Exp./String
•	ASTA/PhosPep/ELM format		I	
9BT23 QAAGAAQATPSHDAKGGGSS1 53667	SKDGLFRVDKGAGNNPEFEETRRV VQRSKSFSLRAQVKETCAACQKTV PVCASCGQRIYDGQYLQALNADWF	YPMERLVADKLIFHNSCFCCKHC	HTKLSLGSYAALHGEFYCKPHFQC QLFCKKDYWARYGESCHGCSEQ	QLFKSKGNYDEGFGRKQHKEL

References

- Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, et al. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127: 635-648.
- 2. Ubersax JA, Ferrell JE, Jr. (2007) Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol 8: 530-541.
- 3. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, et al. (2005) Global analysis of protein phosphorylation in yeast. Nature 438: 679-684.
- 4. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298: 1912-1934.
- 5. Ptacek J, Snyder M (2006) Charging it up: global analysis of protein phosphorylation. Trends Genet 22: 545-554.
- Kobe B, Kampmann T, Forwood JK, Listwan P, Brinkworth RI (2005) Substrate specificity of protein kinases and computational prediction of substrates. Biochim Biophys Acta 1754: 200-209.
- 7. Hjerrild M, Gammeltoft S (2006) Phosphoproteomics toolbox: computational biology, protein chemistry and mass spectrometry. FEBS Lett 580: 4764-4770.
- Kreegipuu A, Blom N, Brunak S, Jarv J (1998) Statistical analysis of protein kinase specificity determinants. FEBS Lett 430: 45-50.
- Songyang Z, Lu KP, Kwon YT, Tsai LH, Filhol O, et al. (1996) A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. Mol Cell Biol 16: 6486-6493.
- 10. Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, et al. (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nat Biotechnol 19: 348-353.
- Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jorgensen C, et al. (2007) Systematic discovery of in vivo phosphorylation networks. Cell 129: 1415-1426.
- Linding R, Jensen LJ, Pasculescu A, Olhovsky M, Colwill K, et al. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Res 36: D695-699.
- 13. Biondi RM, Nebreda AR (2003) Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. Biochem J 372: 1-13.
- Holland PM, Cooper JA (1999) Protein modification: docking sites for kinases. Curr Biol 9: R329-331.
- Tan CS, Linding R (2009) Experimental and computational tools useful for (re)construction of dynamic kinase-substrate networks. Proteomics 9: 5233-5242.
- 16. Erxleben C, Liao Y, Gentile S, Chin D, Gomez-Alegria C, et al. (2006)

Cyclosporin and Timothy syndrome increase mode 2 gating of CaV1.2 calcium channels through aberrant phosphorylation of S6 helices. Proc Natl Acad Sci U S A 103: 3932-3937.

- 17. Gentile S, Martin N, Scappini E, Williams J, Erxleben C, et al. (2008) The human ERG1 channel polymorphism, K897T, creates a phosphorylation site that inhibits channel activity. Proc Natl Acad Sci U S A 105: 14704-14708.
- Ren J, Jiang C, Gao X, Liu Z, Yuan Z, et al. (2010) PhosSNP for systematic analysis of genetic polymorphisms that influence protein phosphorylation. Mol Cell Proteomics 9: 623-634.
- 19. Radivojac P, Baenziger PH, Kann MG, Mort ME, Hahn MW, et al. (2008) Gain and loss of phosphorylation sites in human cancer. Bioinformatics 24: i241-247.
- 20. Diella F, Gould CM, Chica C, Via A, Gibson TJ (2008) Phospho.ELM: a database of phosphorylation sites--update 2008. Nucleic Acids Res 36: D240-244.
- 21. Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci U S A 104: 1488-1493.
- 22. Grimsrud PA, Swaney DL, Wenger CD, Beauchene NA, Coon JJ (2010) Phosphoproteomics for the masses. ACS Chem Biol 5: 105-119.
- Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, et al. (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science 316: 1160-1166.
- Song C, M, Han G, 24. Ye Jiang Х, Wang F, et al. (2010)Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Anal Chem 82: 53-56.
- 25. Hu ZZ, Narayanaswamy M, Ravikumar KE, Vijay-Shanker K, Wu CH (2005) Literature mining and database annotation of protein phosphorylation using a rule-based system. Bioinformatics 21: 2759-2765.
- 26. Xue Y, Ren J, Gao X, Jin C, Wen L, et al. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol Cell Proteomics 7: 1598-1608.

Release Note

- 1. May. 26th, 2010, the alpha version of GPS-PNC was constructed for testing.
- 2. Oct. 28th, 2010, the beta version of GPS-PNC software was released for testing and debugging.
- 3. Dec. 16th, 2010, the stand-alone packages of GPS-PNC 1.0 were released.
- 4. Jul. 23rd, 2011, some bugs were fixed in GPS-PNC 1.0.
- 5. Jan. 3rd, 2012, the GPS-PNC 1.0 was renamed in EPNdb 1.0, while the main program was re-designed as iGPS 1.0. The EPNdb 1.0 was a plugin of iGPS 1.0. The stand-alone packages of iGPS 1.0 were released.
- 6. Aug. 27th, 2012, iGPS 1.0.1 was released. The output format was changed from FASTA to TAB.