PTM Bioinformatics

Yu Xue
Wuhan, China
Huazhong University of Science and Technology
Post-translational modification

- POST: after RNA translated to protein
- Covalent modification: creation or disruption of covalent bond
- Side or main chain of amino acid
- The PTM code: writer-eraser-reader system
- ~610 type of PTMs
  - Phosphorylation (S/T/Y)
  - Acetylation, Ubiquitination, Propionylation (K)

http://www.uniprot.org/docs/ptmlist
PTM bioinformatics in China

- Methods for predicting PTM sites
- Databases of PTMs
- PTM proteome-based analysis
- Tools for PTM analysis
- **Future: PTM function prediction & validation**

Post-translational modification (PTM) bioinformatics in China: progresses and perspectives

Zexian Liu¹, Yudong Cai², Xuejiang Guo³, Ao Li⁴, Tingting Li⁵, Jianding Qiu⁶, Jian Ren⁷, Shaoping Shi⁸, Jiangning Song⁹, Minghui Wang⁴, Lu Xie¹⁰, Yu Xue¹, Ziding Zhang¹¹, Xingming Zhao¹²

http://www.chinagene.cn/CN/abstract/abstract21444.shtml
Phosphorylation

- The most important and well-studied PTM
- Reversibility and regulation:
  - **Protein kinase**: Phosphorylation - *writer*
  - **Phosphatase**: De-phosphorylation - *eraser*
  - **Phospho-binding domain (PDB)**: interacts with phospho-sites - *reader*

The Nobel Prize in Physiology or Medicine 1992

Edmond H. Fischer  Edwin G. Krebs
Known data: 1855 kinases & 347 phosphatases
EKPD: 50,433 kinases and 11,296 phosphatases in 84 eukaryotic species
iEKPD 2.0: 148 species & phospho-binding proteins

In eukaryotes, phosphorylation-dependent signaling networks are, to a large extent, determined by the combined actions of protein kinases, protein phosphatases and phosphoprotein-binding domains (PPBDs) (Lim et al., 2010). Protein kinases are a type of well-understood enzyme which modifies other proteins by chemically adding phosphate groups to them (phosphorylation). Phosphorylation has been found to be involved in a variety of cellular processes, including metabolism (Vicili and Andreelli, 2011), transcription, cell cycle progression (Moniz et al., 2011), cytoskeleton rearrangement and cell movement (Huang et al., 2009), cell apoptosis (Wang, 2000), and differentiation (Taylor et al., 2011). Contrary to phosphorylation, dephosphorylation is catalyzed by protein phosphatases through a way of removing a phosphate group from its substrate. Protein phosphatases contain two groups, PSP and PTP, which are responsible for the ser/thr dephosphorylation and tyr dephosphorylation, respectively. Like protein kinases, protein phosphatases also play an important role in many cellular processes, including proliferation (Zhang et al., 2012), differentiation, cell adhesion (Bessette et al., 2008), motility and cell death (Gallego et al., 2005). More over, recent studies discovered a few modular domains that particularly recognize pThr/pSer- or pThr-containing sequences, such as the breast-cancer-associated protein BRCA1 C-terminal (BRCT) repeats, SH2 domain and forhead-associated domain. These PPBD-containing proteins play a pivotal role in connecting the kinases and other effector molecules.

In this work, we have collected 1803 protein kinases, 383 protein phosphatases and 411 PPBDs-containing proteins from the scientific literature and various public databases. The data are further classified into 33 families for protein kinase, 148 families for protein kinase and 21 families for PPBD-containing proteins, respectively. To computationally detect more proteins in eukaryotes, we constructed hidden Markov model (HMM) profiles for these families. For families without HMM, we contained 150,000 unique prote

http://iekpd.biocuckoo.org/

Wang et al., NAR, 2014, 42:D496-502
Xu et al., unpublished
The classifications

- **Kinase**: 10 groups with 149 families
- **Phosphatase**: 10 groups with 33 families
Animals vs. plants

- **Kinase:** 467 vs. 1,450
- **Phosphatase:** 144 vs. 192

8.2% vs. 60.2%
dbPPT

- Manual curation: 82,175 p-sites in 31,012 proteins for 20 plant species

As one of the most important and ubiquitous post-translational modifications (PTMs) in plants, the reversible phosphorylation catalyzed by protein kinases is involved in regulating a wide range of biological processes such as cellular metabolism, signal transduction and environmental response (Ranjeva et al., 1987, Mundy et al., 2002, Kusakina et al., 2012). Recently with the development of phosphopeptide enrichment techniques and high-throughput mass spectrometry (HTP-MS) technology, large scale phosphoproteomics data in plants have been emerged at a quick pace, stimulating the creation of multiple phosphorylation databases to store and organize them. However, a more comprehensive phosphorylation database of plants is still necessary and useful. In this work, we develop a database of dbPPT 1.0 (database of Phospho-sites in PlantTs), which contains large-scale phosphorylation sites (p-sites) based on mass spectrometry across 20 plant species. Most of the datasets in dbPPT 1.0 were manually curated from literature published before July, 2014. Moreover, we also integrated other phosphorylation resource into this database including PhosPhAt (Durek et al., 2010) and P5DB (Yao et al., 2014). Currently, dbPPT contains 82,175 p-sites in 31,012 proteins for 20 plant species.

Cheng et al., Database (Oxford), 2014, bau121
dbPSP

- 7,391 p-sites in 3,750 prokaryotic proteins
dbPAF

- 483,001 p-sites of 54,148 proteins for human, animals and fungi

The dbPAF (database of Phospho-sites in Animals and Fungi) is an online data resource specifically designed for protein phosphorylation in seven eukaryotic species, including *H. sapiens*, *M. musculus*, *R. norvegicus*, *D. melanogaster*, *C. elegans*, *S. pombe* and *S. cerevisiae*. From the scientific literature, we collected 294,370 non-redundant phosphorylation sites of 40,432 proteins. We also integrated known phosphorylation sites from a number of public databases, such as PhosphoELM (Diella, et al., 2004; Diella, et al., 2008), dbPTM (Huang, et al., 2015; Lee, et al., 2008), PHOSIDA (Gnad, et al., 2011; Olsen, et al., 2006), PhosphositePlus (Hornbeck, et al., 2004; Hornbeck, et al., 2015), PhosphoPep (Bodenmiller, et al., 2008; Bodenmiller, et al., 2007), PhosphoGRID (Sadowski, et al., 2013; Stark, et al., 2010), SysPTM (Li, et al., 2009; Li, et al., 2014), HPRD (Goel, et al., 2012) and UniProt (The UniProt Consortium, 2015). In total, dbPAF 1.0 contained 483,001 known phosphorylation sites of 54,148 protein substrates, as a comprehensive data resource for human, animals and fungi.

**Overview**

Please search the dbPAF database with one or multiple keywords to find the related information.
CPLM & PLMD

- CPLM 2.0: 12 types of protein lysine modifications
- PLMD 3.0: 20 lysine modifications

Liu et al., 2014, NAR, 42, D531-6
Xu et al., J Genet Genomics, 2017, 44, 243-250
Submit to JGG

- *Journal of Genetics and Genomics*
- Annual database/web server issue, IF: 4.051
- Article & Letter
- Guest Editors: Xiujie Wang & Yu Xue
- Submission: Early of December, 2017
- Publication: May, 2018
- E-mail: xueyu@hust.edu.cn
PTMomics

- Large-scale detection of \textit{in vivo} PTM substrates, sites and motifs
  - Mass spectrometry (MS)
  - Protein, peptide, and PTM chips

- Current progress:
  - \(~500,000\) phosphorylation sites
  - \(~140,000\) ubiquitination sites
  - \(~60,000\) acetylation sites

- \textbf{Challenge}: What PTM Bioinformatics can do?
Autophagy

- **Macroautophagy**, microautophagy, chaperon-mediated autophagy (CMA)
- 1963, C de Duve, “self-eating” in Greek
- 1993, Atg1 in yeast
- 41 core ATG genes
- ~20 conserved in human

Xie et al., Autophagy, 2015, 11, 28-45
Neuronal autophagy

- In Alzheimer’s & Parkinson’s diseases
  - Proteins accumulate in central nervous system
  - Defective autophagy in patient brains

- Enhanced autophagy
  - Neuroprotective by promoting the clearance of disease-associated aggregates

- Small-molecule autophagy enhancers
  - *Uncaria rhynochophylla* (Gouteng,钩藤)
Neuroprotective alkaloids in Gouteng

- Corynoxine (柯诺辛碱) & corynoxine B (柯诺辛B)
  - Same molecular formula, different conformation
  - Induce autophagy in different way

Question:
- Find key regulators in neuronal autophagy
- Distinguish two compounds

Lu et al., Autophagy, 2012, 8, 98-108
Chen et al., J Neuroimmune Pharmacol., 2014, 9, 380-7
Experimental procedure

- **N2a**: a mouse neuroblastoma cell line

  ![Diagram of experimental procedure]

  **Tripsin digestion**

  Treatment/Control ratio: >1 or < 1
Phosphoproteomics profiling

- Quantification: 2,317 proteins and 5,555 unique p-sites
- GSEA-based enrichment analysis
  - GO biological processes
- Limited difference in the distribution of p-sites
‘in vivo’ GPS

- ‘Kiss-then-farewell’ model: the interaction of kinase-substrate
- iGPS:
  - GPS algorithm
  - Protein-protein interaction
  - The phosphoproteomic data
  - Much better than NetworKIN

<table>
<thead>
<tr>
<th>PK clusters</th>
<th>NetworKIN Ac</th>
<th>NetworKIN Sn</th>
<th>NetworKIN Sp</th>
<th>NetworKIN MCC</th>
<th>iGPS Ac</th>
<th>iGPS Sn</th>
<th>iGPS Sp</th>
<th>iGPS MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetworKIN 1.1</td>
<td>98.87%</td>
<td>58.89%</td>
<td>99.44%</td>
<td>0.5865</td>
<td>98.81%</td>
<td>52.22%</td>
<td>99.47%</td>
<td>0.5445</td>
</tr>
<tr>
<td>AGC/AKT</td>
<td>92.24%</td>
<td>32.20%</td>
<td>93.28%</td>
<td>0.1277</td>
<td>92.67%</td>
<td>57.56%</td>
<td>93.28%</td>
<td>0.2481</td>
</tr>
<tr>
<td>AGC/PKA</td>
<td>97.50%</td>
<td>77.42%</td>
<td>97.83%</td>
<td>0.5204</td>
<td>97.47%</td>
<td>75.27%</td>
<td>97.83%</td>
<td>0.5080</td>
</tr>
<tr>
<td>Atypical/PIKK/ATM</td>
<td>98.43%</td>
<td>5.85%</td>
<td>99.91%</td>
<td>0.1671</td>
<td>98.51%</td>
<td>10.64%</td>
<td>99.91%</td>
<td>0.2580</td>
</tr>
<tr>
<td>CAMK/CAMK2</td>
<td>93.45%</td>
<td>65.34%</td>
<td>94.06%</td>
<td>0.3312</td>
<td>93.60%</td>
<td>71.12%</td>
<td>94.09%</td>
<td>0.3614</td>
</tr>
<tr>
<td>CMGC/MAPK</td>
<td>94.58%</td>
<td>46.67%</td>
<td>95.69%</td>
<td>0.2825</td>
<td>94.75%</td>
<td>54.04%</td>
<td>95.69%</td>
<td>0.3268</td>
</tr>
<tr>
<td>CMGC/CDK/CDC2</td>
<td>90.17%</td>
<td>55.49%</td>
<td>91.14%</td>
<td>0.2511</td>
<td>90.25%</td>
<td>54.85%</td>
<td>91.24%</td>
<td>0.2495</td>
</tr>
<tr>
<td>Other/CK2</td>
<td>85.26%</td>
<td>12.16%</td>
<td>96.97%</td>
<td>0.1554</td>
<td>91.42%</td>
<td>56.76%</td>
<td>96.97%</td>
<td>0.6059</td>
</tr>
<tr>
<td>TK/EGFR</td>
<td>90.04%</td>
<td>23.97%</td>
<td>95.40%</td>
<td>0.2141</td>
<td>90.69%</td>
<td>32.64%</td>
<td>95.40%</td>
<td>0.2956</td>
</tr>
<tr>
<td>TK/Src</td>
<td>90.04%</td>
<td>23.97%</td>
<td>95.40%</td>
<td>0.2141</td>
<td>90.69%</td>
<td>32.64%</td>
<td>95.40%</td>
<td>0.2956</td>
</tr>
</tbody>
</table>

Song et al., MCP, 2012, 11, 1070-1083
Neuronal autophagy phosphorylation network

- iGPS: Cory- & Cory B-regulated phosphorylation networks
- Single kinase network
  - Down-regulated network & up-regulated network
iKAP algorithm

For Kinase \( i \), statistically test whether it prefer to be involved in up-regulated (high activity) or down-regulated (low activity) networks

\[ KA: \text{kinase activity}; \ KS: \text{Treatment/Control (T/C) ratio} \]

Up-regulated network

\[ T/C > 1, \ KA_{up}(i) = \sum_{j=1}^{m} \text{int}(KS_{ij}) \]

Down-regulated network

\[ T/C < 1, \ KA_{Down}(i) = \sum_{j=1}^{n} \text{int}\left(\frac{1}{KS_{ij}}\right) \]

Yates’ chi-squared test

\[ KA_{up} = \sum_{i=1}^{k} KA_{up}(i), \ KA_{down} = \sum_{i=1}^{l} KA_{down}(i) \]
Differentially activated kinases

- Up-regulated: 28 (Cory) & 28 (Cory B)
- Down-regulated: 51 (Cory) & 27 (Cory B)
THANATOS database

- **The Autophagy, Necrosis, Apoptosis OrchestratorS**
  - 144,153 proteins in 148 eukaryotes
  - Autophagy: 119 mouse kinases

- THANATOS filter
  - Up-regulated: 6 (Cory) & 4 (Cory B)
  - Down-regulated: 12 (Cory) & 7 (Cory B)

---

**a. The THANATOS database**

THANATOS (The Apoptosis, Necrosis, Autophagy OrchestratorS) is a resource being developed by the CUCKOO Workgroup at the Huazhong University of Science and Technology (Wuhan, Hubei, China). THANATOS is still under development (Y. Xue, personal communication) and it is focused on the integration of sequence data related to the main mechanisms leading to programmed cell death in eukaryotes. A simple web interface assists in data retrieval, using keyword searches, browsing by species and cell death type, performing BLAST searches with user-defined sequences, and by requesting the display of orthologs among predefined species. A Java application is also available to download for standalone usage of the THANATOS resource. The THANATOS database is publicly available online at the URL http://thanatos.biocuckoo.org/.
Validation

- Cory: down-regulates p70S6K and up-regulates MEK2 & PLK1
MEK2 & PLK1

- Silencing MEK2 but not PLK1 decreases LC3 II
- Silencing MEK2 & PLK1 both increase p62

LC3 II ↑: autophagy inhibition & activation
p62 ↓: autophagic flux ↑
MEK2 & PLK1 activation

- Inhibitors: U0126 (MEK2) & BI2356 (PLK1)
- The inhibitions of MEK2 and PLK1 both increase p62 and block autophagic flux
MEK2 & PLK1 in neuronal autophagy

- Alzheimer's disease: APP (β-amyloid precursor protein) & CTF β
- Parkinson’s disease: α-synuclein (α-syn)
- The inhibition of MEK2 or PLK1 diminishes the clearance of disease-associated proteins by Cory
- The activation of MEK2 & PLK1: neuroprotective
Summarization

- **PTM databases**
  - Phosphorylation: EKPD, dbPPT, dbPSP, dbPAF
  - Lysine modifications: CPLM, PLMD

- **Functional PTMs**
  - Functional protein kinases
  - iKAP: a network-based algorithm
  - Cory inhibits p70S6K and activates MEK2 & PLK1
  - Inhibition of MEK2 & PLK1 block autophagic flux
  - Activation of MEK2 & PLK1 is neuroprotective to clear disease-associated proteins
Acknowledgements

➢ Collaborators

Prof. Jian Ren, SYSU
Prof. Li Yu, Tsinghua
Prof. Min Li, HKBU
Prof. Xiang-Jiao Yang, McGill
Leilei Chen

➢ Students

◆ Zexian Liu,
◆ Tianshun Gao
◆ Han Cheng
◆ Yongbo Wang
◆ Wankun Deng
◆ Ying Zhang
◆ Jiaqi Zhou
◆ Shaofeng Lin
◆ Yang Xu
◆ Shuang Zhang

Prof. Zexian Liu, SYSU
Thanks!

Any questions?
Illustrator for Biological Sequences (IBS)

The 3.8 Å resolution cryo-EM structure of Zika virus

http://ibs.biocuckoo.org
Heatmap Illustrator (HemI)

- http://hemi.biocuckoo.org