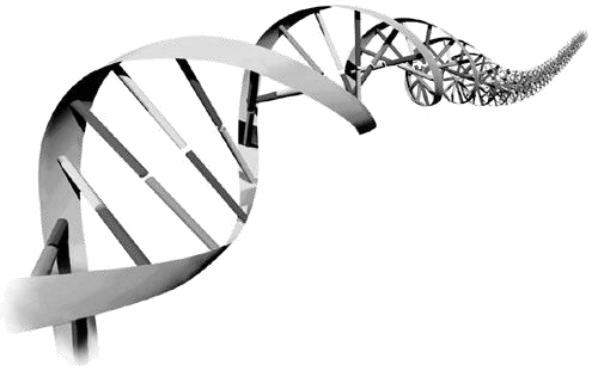
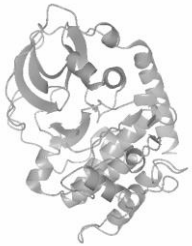




PTM Bioinformatics

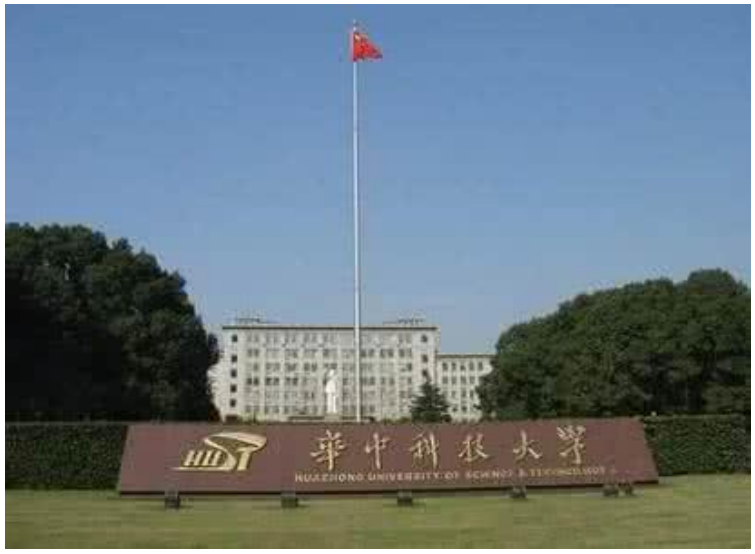
Yu Xue

Huazhong University of Science and Technology

Wuhan, China

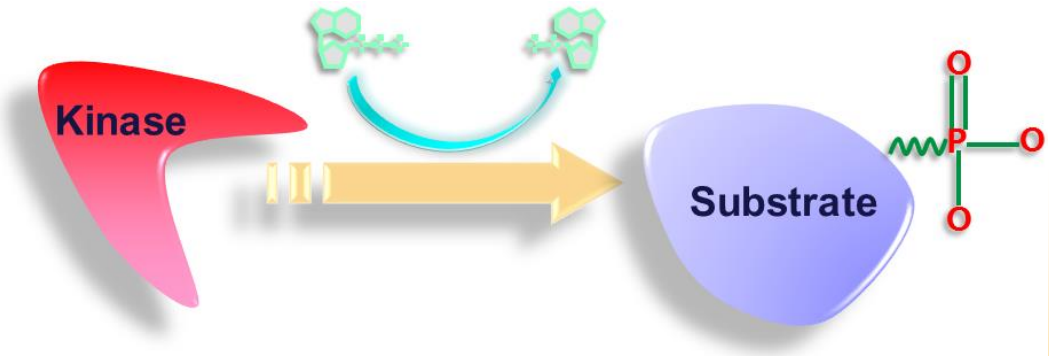

巴甫洛达尔
乌兰乌德
赤塔
共青城
卡拉干达
乌斯季卡缅诺戈尔斯克
哈巴罗夫斯克（伯力）
乌兰巴托
哈尔滨
乌鲁木齐
长春
阿拉木图
札幌
沈阳
清津
函馆
呼和浩特
北京
咸兴
天津
平壤
日本海
渤海
银川
石家庄
仙台
太原
首尔
济南
西宁
兰州
大田
金泽
东京
光州
斯利那加
西安
郑州
黄海
广岛
大阪
长崎
费萨拉巴德
合肥
南京
鹿儿岛
上海
成都
武汉
杭州
拉萨
重庆
南昌
巴雷利
新德里
加德满都
廷布
长沙
阿格拉
阿杰梅尔
坎普尔
贵阳
福州
东海
古瓦哈蒂
安拉阿巴德
钓鱼岛
赤尾屿
昆明
台北
乌代布尔
达卡
广州
印多尔
贾巴尔普尔
兰契
南宁
香港
加尔各答
曼德勒
澳门
河内
苏拉特
那格浦尔
东沙群岛
纳西克
克塔克
内比都
海口
孟买
万象

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

遗传 Hereditas (Beijing) 2015 年 7 月, 37(7): 621－634
www.chinagene.cn

综 述

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

Zexian Liu[1], Yudong Cai[2], Xuejiang Guo[3], Ao Li[4], Tingting Li[5], Jianding Qiu[6], Jian Ren[7], Shaoping Shi[8], Jiangning Song[9], Minghui Wang[4], Lu Xie[10], Yu Xue[1], Ziding Zhang[11], Xingming Zhao[12]

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*

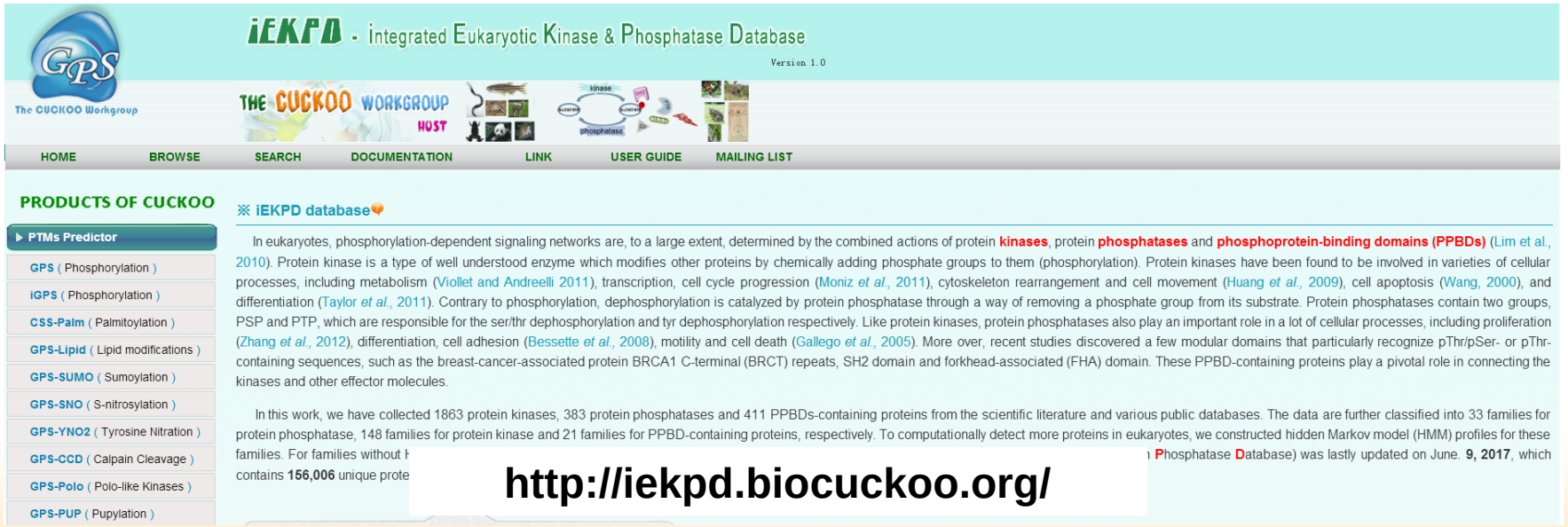
Edmond H. Fischer

Edwin G. Krebs

The Nobel Prize in Physiology or Medicine 1992

EKPD: Eukaryotic Kinase and Phosphatase Database

- **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**



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**

AGC	
Akt	194.2
DMPK	81.2
GRK	188.5
MAST	86.3
NDR	113.1
PDK1	98.5
PKA	153.2
PKC	164.6
PKG	190.5
PKN	159.0
RSK	101.2
RSKL	81.5
RSKR	139.7
SGK	152.7
YANK	140.6
Unclassified	38.9

CMGC	
CDK	61.1
CDKL	110.8
CLK	91.2
DYRK	97.7
GSK	124.1
MAPK	61.5
RCK	92.0
SRPK	120.9
Unclassified	109.9

CAMK	
CAMK1	125.5
CAMK2	116.4
CAMKL	89.9
CASK	183.8
DAPK	85.6
DCAMKL	150.4
MAPKAPK	112.7
MLCK	139.6
PHK	140.0
PIM	149.7
PKD	101.6
PSK	284.6
RAD53	106.7
RSKb	78.9
TSSK	125.4
Trbl	151.2
Trio	68.8
Unclassified	102.8

STE	
STE11	113.4
STE20	42.2
STE7	75.6
Unclassified	117.2

CK1	
CK1	98.2
Dual	407.8
TTBK	258.1
TTBKL	164.1
VRK	112.2
Worm10	434.6
Worm6	76.0
Worm7	203.5
Worm8	524.7
Worm9	535.4
Unclassified	63.4

TKL	
IRAK	88.6
LISK	75.8
LRRK	118.2
MLK	74.9
RAF	102.9
RIPK	113.7
STKR	89.5
Unclassified	136.5

Atypical	
ABC1	71.4
Alpha	36.7
PDHK	120.3
PIKK	110.4
RIO	71.5

TK	
Abl	367.5
Ack	97.8
Alk	80.8
Axl	159.8
CCK4	196.2
Csk	144.6
DDR	60.8
EGFR	101.6
Eph	120.1
FAK	190.0
FGFR	192.0
Fer	73.6
InsR	232.2
Jak	68.0
KIN16	282.8
KIN6	414.3
Lmr	127.9
Met	95.0
Musk	138.1
PDGFR	139.2
Ret	194.5
Ror	321.9
Ryk	89.5
Sev	187.7
Src	115.7
Syk	216.9
Tec	116.1
Tie	346.4
Trk	70.2
VEGFR	184.4
Unclassified	110.1

RGC	
RGC	70.8

Other	
Aur	159.2
BUB	87.8
Bud32	93.1
CAMKK	138.3
CDC7	80.3
HAL	185.7
Haspin	124.7
IKK	135.1
IRE	86.2
MOS	118.7
NAK	162.3
NEK	81.4
NKF1	137.0
NKF2	125.0
NKF3	175.9
NKF4	180.9
NKF5	92.3
NRBP	154.5
PEK	107.2
PLK	75.5
RAN	176.4
SCY	79.3
Slob	232.5
TBCK	58.8
TLK	139.9
TOPK	149.2
TTK	161.9
ULK	64.1
VPS15	220.9
WEE	63.6
WNK	83.0
Worm1	581.4
Worm2	492.7
Worm3	639.1
Worm4	413.5
Worm5	436.4
Unclassified	30.4

PK families

PPP	
PP1	169.7
PP2A	120.6
PP2B	79.6
PP4	161.0
PP5	120.9
PP6	168.1
PP7	72.8
Kelch	301.7
SLP	67.4
Unclassified	94.9

Asp Based PTP	
FCP_SCP	30.4
EYA	49.7
MDP1	145.6

Classical PTP	
RPTP	62.1
NRPTP	55.5

DSP	
aDSP	37.2
MKP	69.0
PRL	109.6
CDC14	60.3
SSH	72.7
Myotubularins	74.5
PTEN	65.0

PPM	
PP2C_1	25.6
PP2C_2	38.7
PP2C_3	24.7
PDP	171.0

LMWPTP	
LMWPTP	77.2

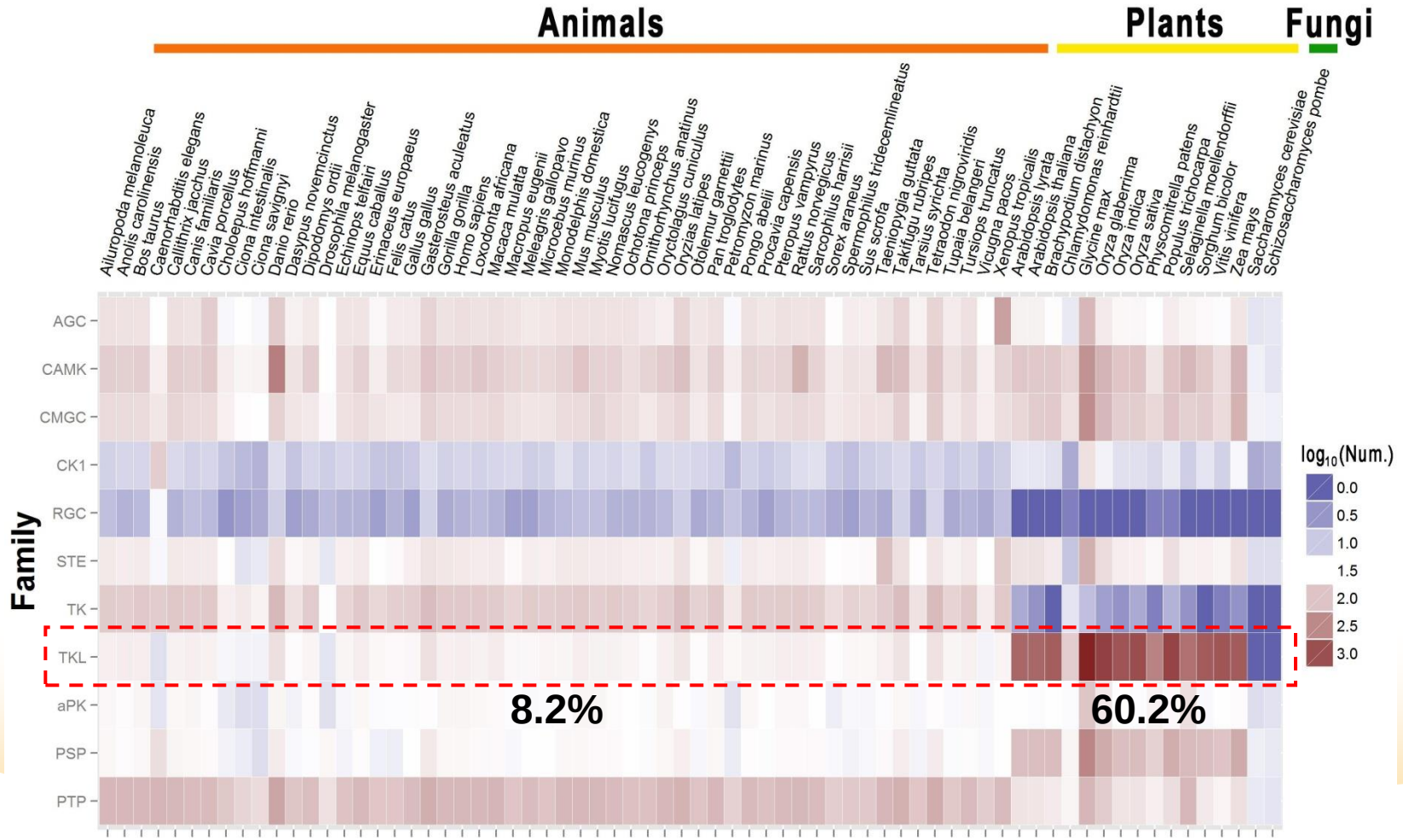
CDC25	
CDC25	50.5

PTPLA	
PTPLA	35.5

PP families

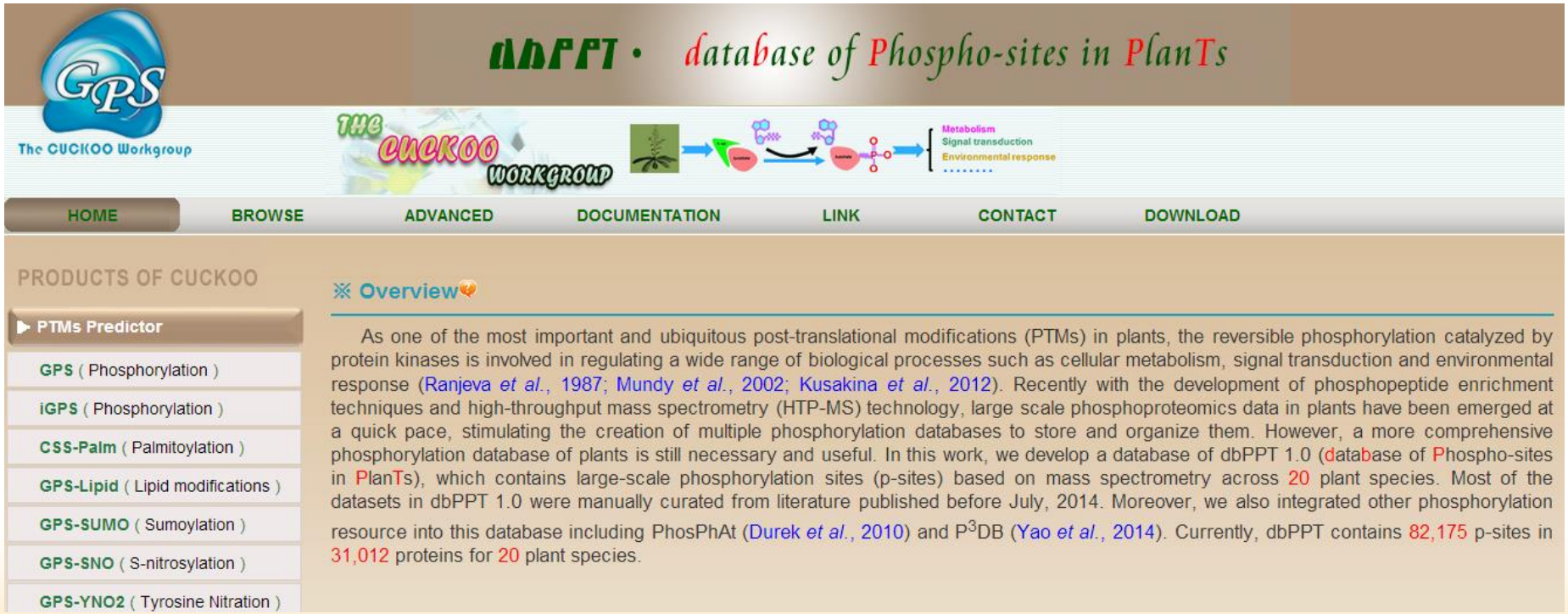
Animals *vs*. plants

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



dbPPT

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



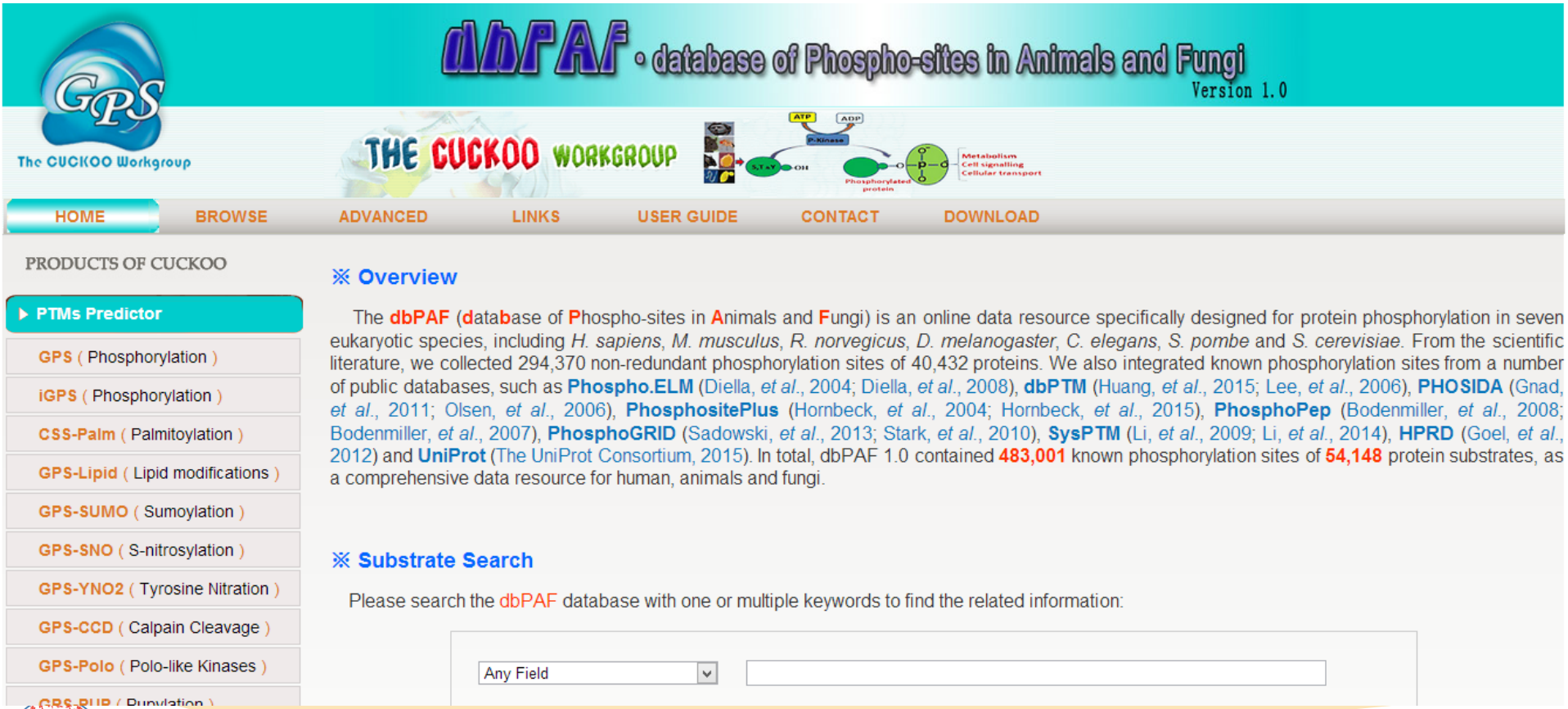
dbPSP

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

Pan et al., Database (Oxford), 2015, bav031

dbPAF

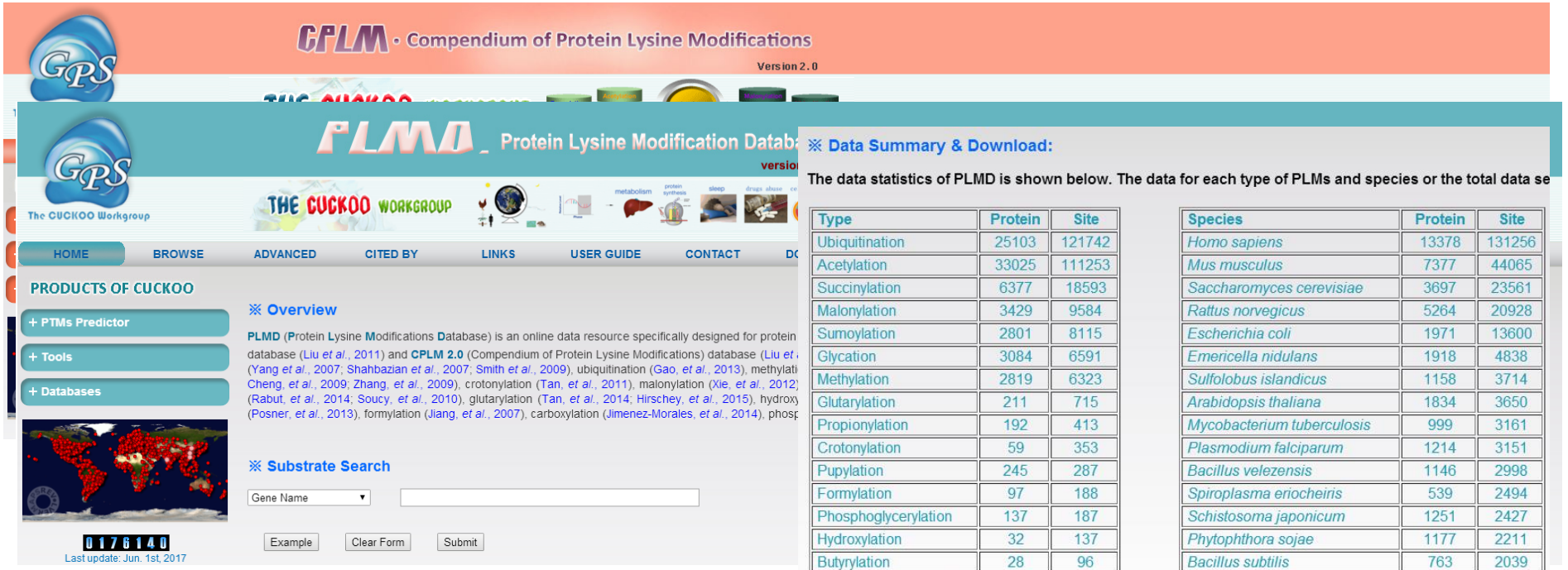
- **483,001 p-sites of 54,148 proteins for human, animals and fungi**



Ullah et al., Sci Rep, 2016, 6, 23534

CPLM & PLMD

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



Type	Protein	Site
Ubiquitination	25103	121742
Acetylation	33025	111253
Succinylation	6377	18593
Malonylation	3429	9584
Sumoylation	2801	8115
Glycation	3084	6591
Methylation	2819	6323
Glutarylation	211	715
Propionylation	192	413
Crotonylation	59	353
Pupylation	245	287
Formylation	97	188
Phosphoglycerylation	137	187
Hydroxylation	32	137
Butyrylation	28	96
2-hydroxyisobutyrylation	15	81
Neddylation	25	50
Carboxylation	36	36
Lipoylation	25	28
Biotinylation	8	8
Total	**53501**	**284780**

Species	Protein	Site
Homo sapiens	13378	131256
Mus musculus	7377	44065
Saccharomyces cerevisiae	3697	23561
Rattus norvegicus	5264	20928
Escherichia coli	1971	13600
Emericella nidulans	1918	4838
Sulfolobus islandicus	1158	3714
Arabidopsis thaliana	1834	3650
Mycobacterium tuberculosis	999	3161
Plasmodium falciparum	1214	3151
Bacillus velezensis	1146	2998
Spiroplasma eriocheiris	539	2494
Schistosoma japonicum	1251	2427
Phytophthora sojae	1177	2211
Bacillus subtilis	763	2039
Drosophila melanogaster	1051	2026
Oryza sativa subsp	913	2018
Vibrio parahaemolyticus	643	1929
Corynebacterium glutamicum	637	1897
Others	6571	12817
Total	**53501**	**284780**

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

Submit to JGG

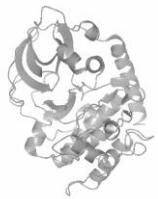
- *Journal of Genetics and Genomics*
- **Annual database/web server issue, IF: 4.051**
- **Article & Letter**
- **Guest Editors: Xiujie Wang & Yu Xue**
- **Submission: Early of December, 2017**
- **Publication: May, 2018**
- **E-mail: xueyu@hust.edu.cn**

Journal of Genetics and Genomics
Volume 44, Issue 5, Pages 223-284 (20 May 2017)
Edited by Xiu-Jie Wang and Yu Xue

Editorial

Bioinformaticians wrestling with the big biomedical data

Yu Xue, Xiu-Jie Wang

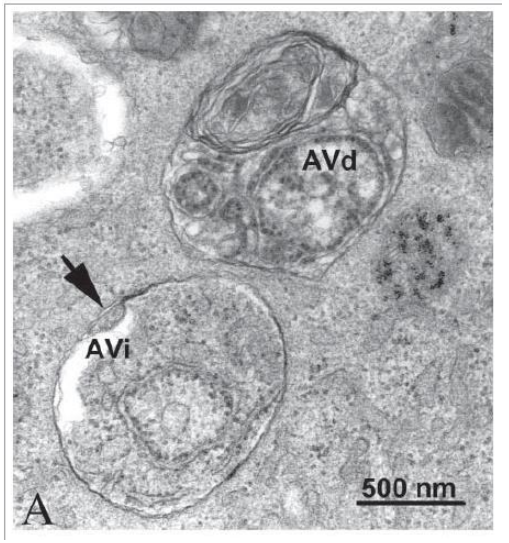


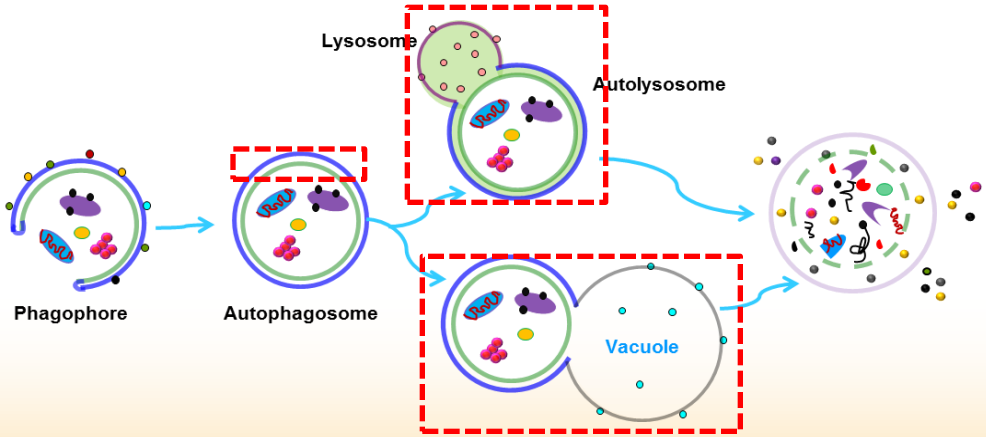
PTMomics

- **Large-scale detection of *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**
- **Challenge: What PTM Bioinformatics can do?**

Autophagy

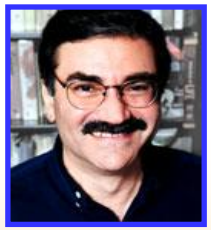
- **Macroautophagy**, microautophagy, chaperon-mediated autophagy (CMA)
- 1963, C de Duve, “self-eating” in Greek
- 1993, Atg1 in yeast
- 41 core ATG genes
- ~20 conserved in human






Christian de Duve


Yoshinori Ohsumi
（大隅良典）


Daniel J. Klionsky

Ohsumi Y, Cell Res., 2014, 24, 9-23
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 rhynchophylla* (Gouteng,钩藤)**

草部·钩藤

作者：李时珍

气味

甘、微寒、无毒。

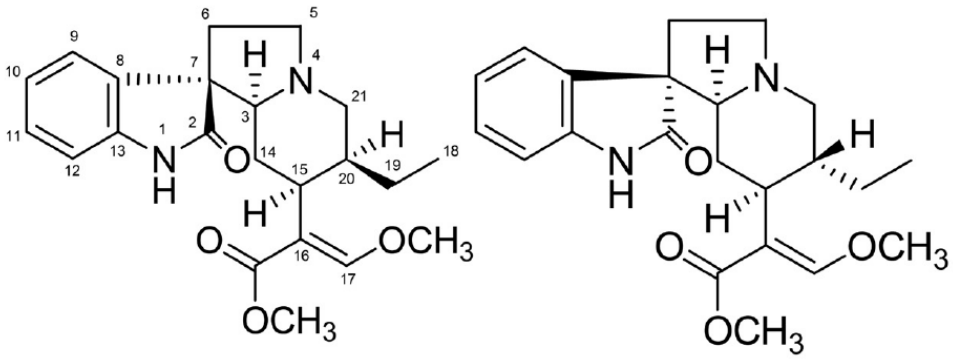
主治

小儿惊热。用钩藤一两、硝石半两，甘草（炙）一分，共研为末。每服半钱，温水服，一天服三次。此方名"延龄散"。

班疹。用钩藤的钩子、紫草茸，等分为末。每服三分或半钱，温酒送下。

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

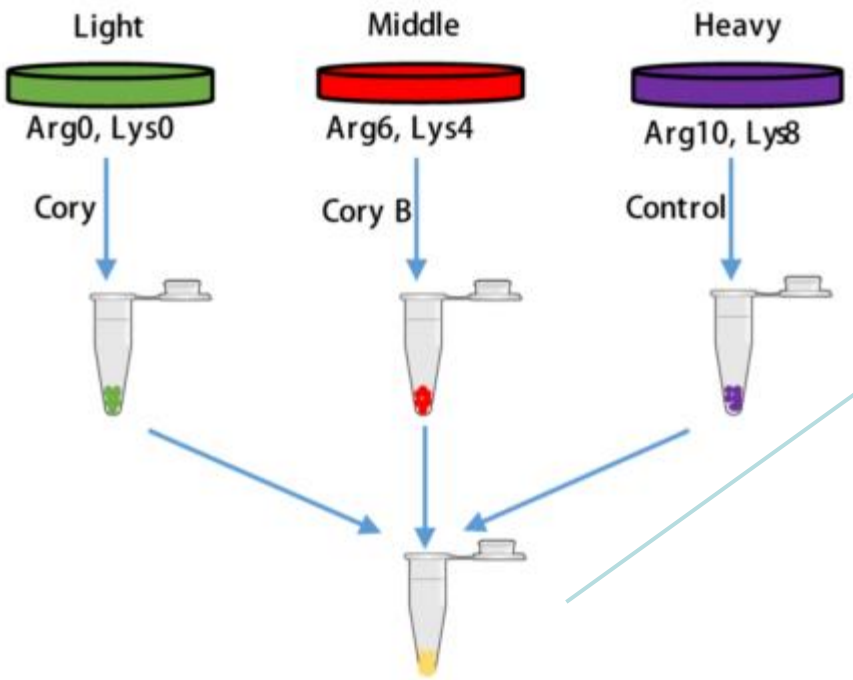


Corynoxine B　　Corynoxine

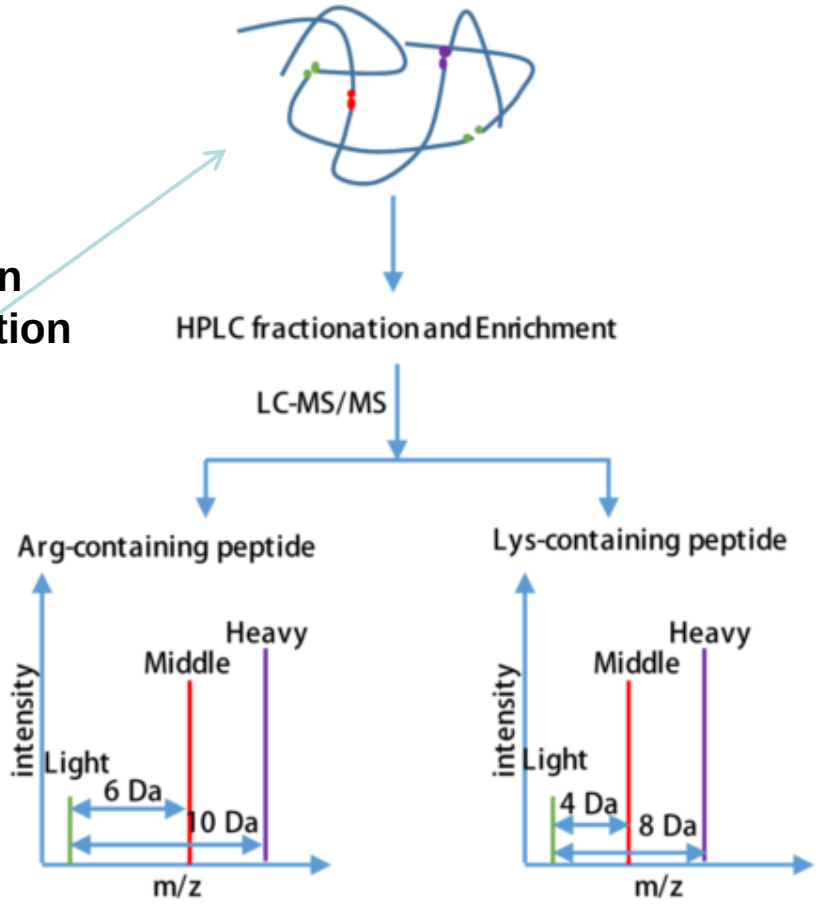
Lu et al., Autophagy, 2012, 8, 98-108
Chen et al., J Neuroimmune Pharmacol., 2014, 9, 380-7

Experimental procedure

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



Tripsin digestion

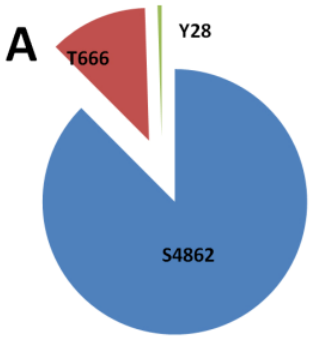


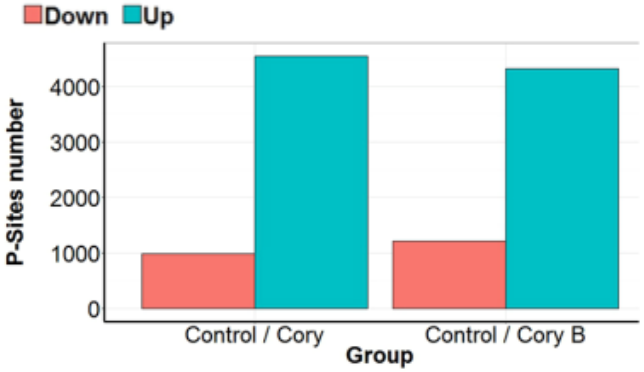
Treatment/Control ratio: >1 or < 1

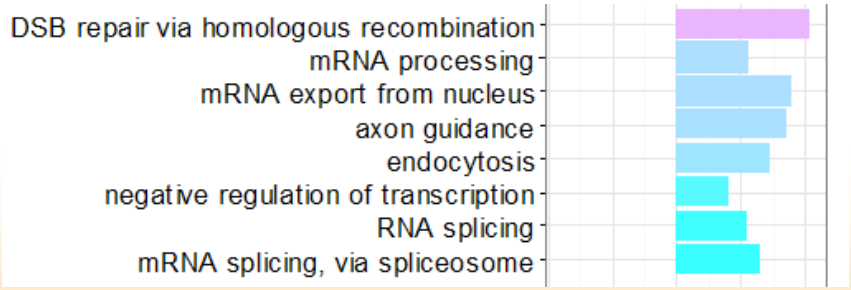
Chen et al., Autophagy, 2017, in press

Phosphoproteomics profiling

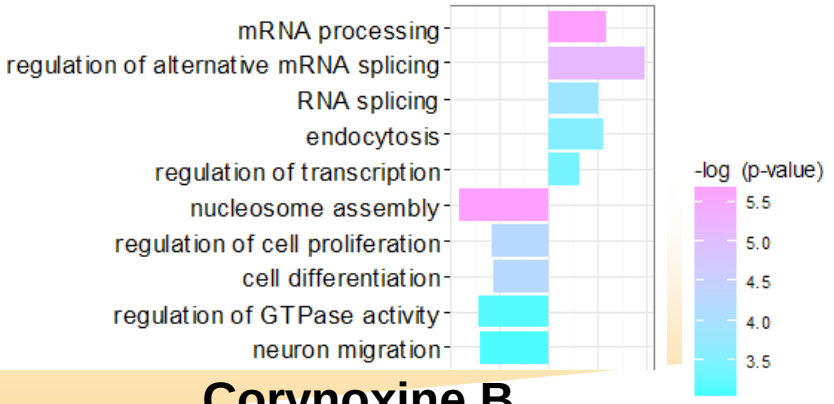
- Quantification: 2,317 proteins and 5,555 unique p-sites
- GSEA-based enrichment analysis
  - ◆GO biological processes
- Limited difference in the distribution of p-sites







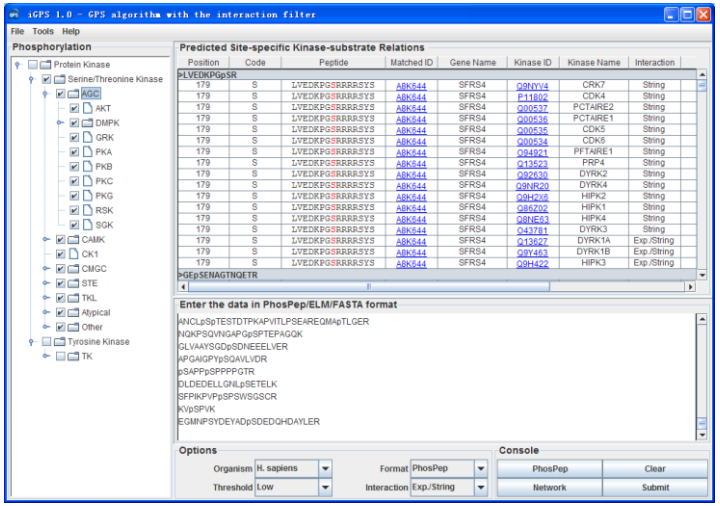
Corynoxine



Corynoxine B

'*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**

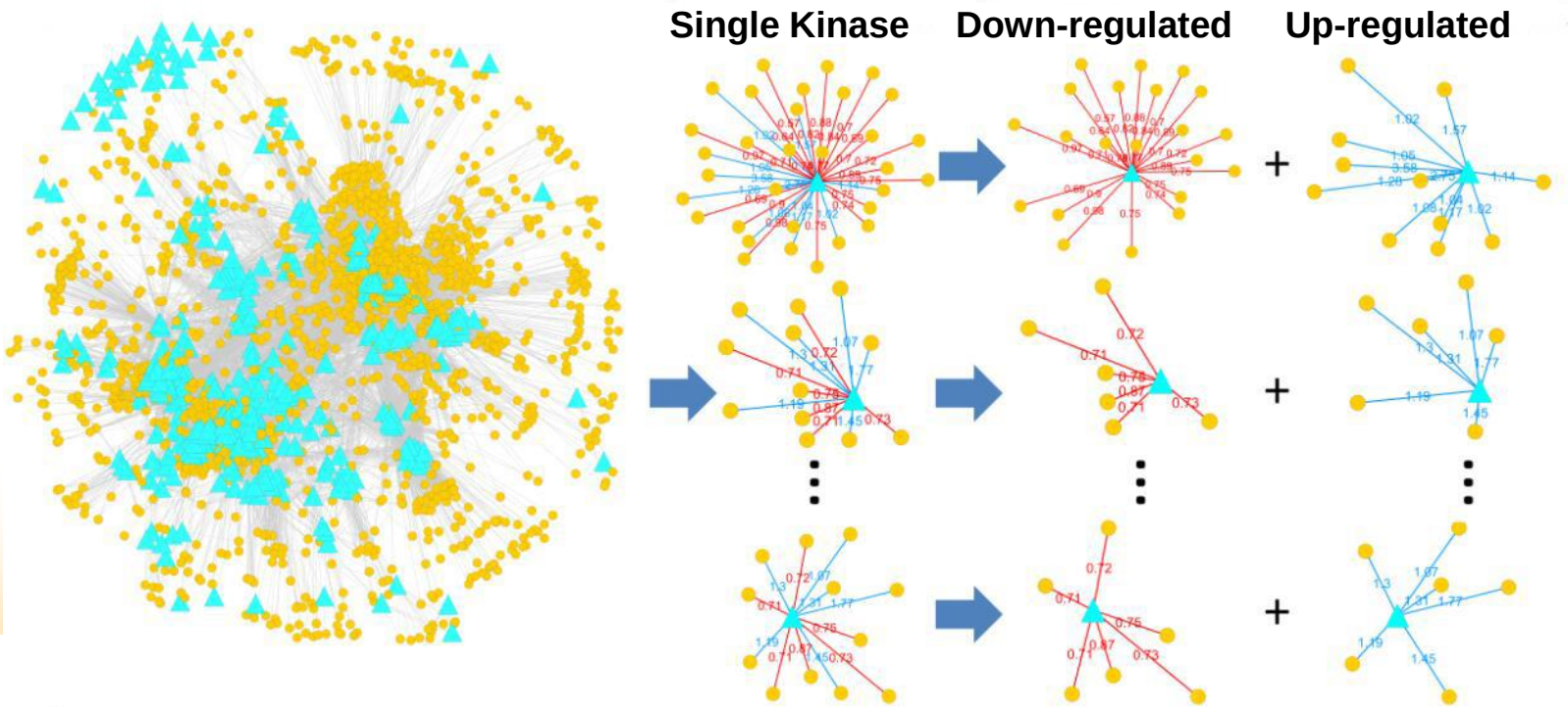


PK clusters	NetworKIN				iGPS			
	Ac	*Sn*	*Sp*	*MCC*	*Ac*	*Sn*	*Sp*	*MCC*
NetworKIN 1.1								
AGC/AKT	**98.87%**	**58.89%**	**99.44%**	**0.5865**	**98.81%**	**52.22%**	**99.47%**	**0.5445**
AGC/PKA	92.24%	32.20%	93.28%	0.1277	92.67%	57.56%	93.28%	0.2481
Atypical/PIKK/ATM	**97.50%**	**77.42%**	**97.83%**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

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

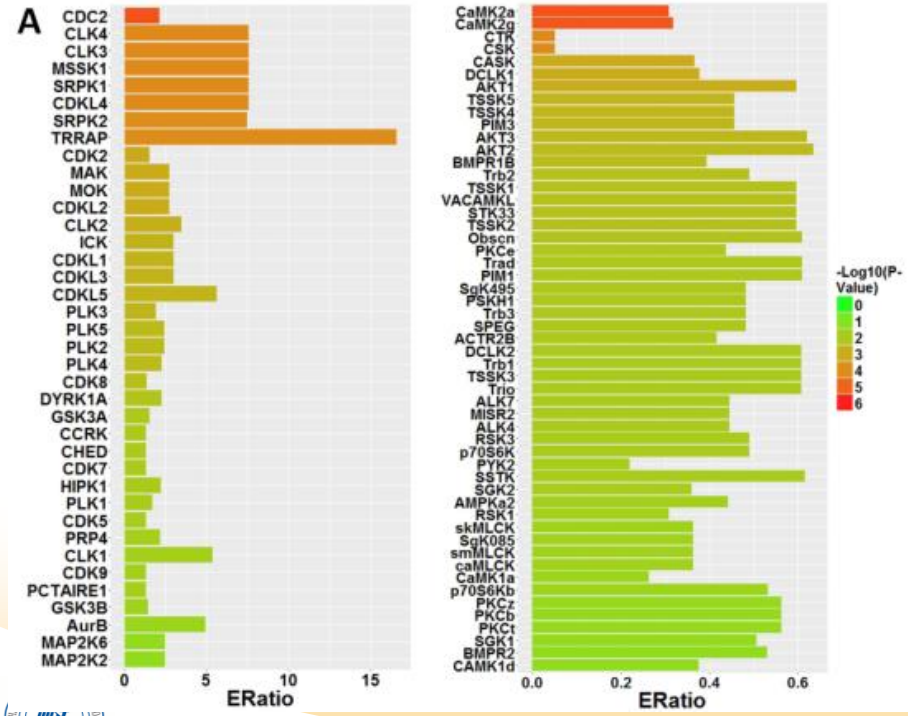


iKAP algorithm

- For Kinase *i*, statistically test whether it prefer to be involved in up-regulated (high activity) or down-regulated (low activity) networks
  - *KA*: kinase activity; *KS*: Treatment/Control (T/C) ratio
- Up-regulated network
  - T/C >1, $KA_{up}(i) = \sum_{j=1}^{m} int(KS_{ij})$
- Down-regulated network
  - T/C < 1, $KA_{Down}(i) = \sum_{j=1}^{n} int(\frac{1}{KS_{ij}})$
- 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)**

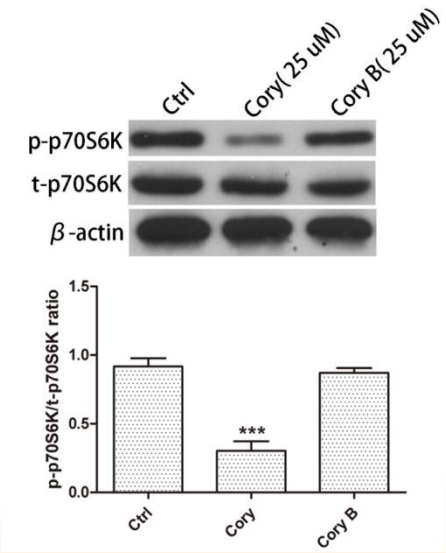
a. *The THANATOS database*

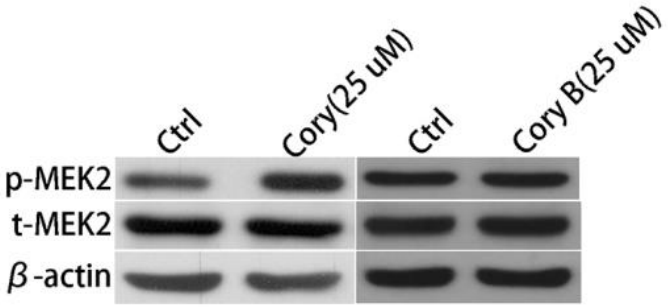
THANATOS (THe Apoptosis, Necrosis, AuTophagy OrchestratorS) is a resource being developed by the CUCKOO Workgroup at the Huazhong University of Science and Technology (Wuhan, Hubei,China). THANATOS is still under development (Y. Xue, personal communication) and it is focused on the integration of sequence 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/.

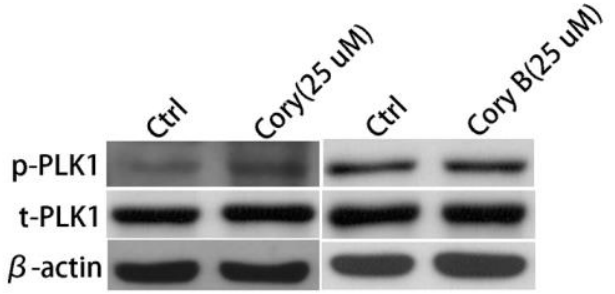
Klionsky et al., Autophagy, 2016, 12, 1-222

Validation

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

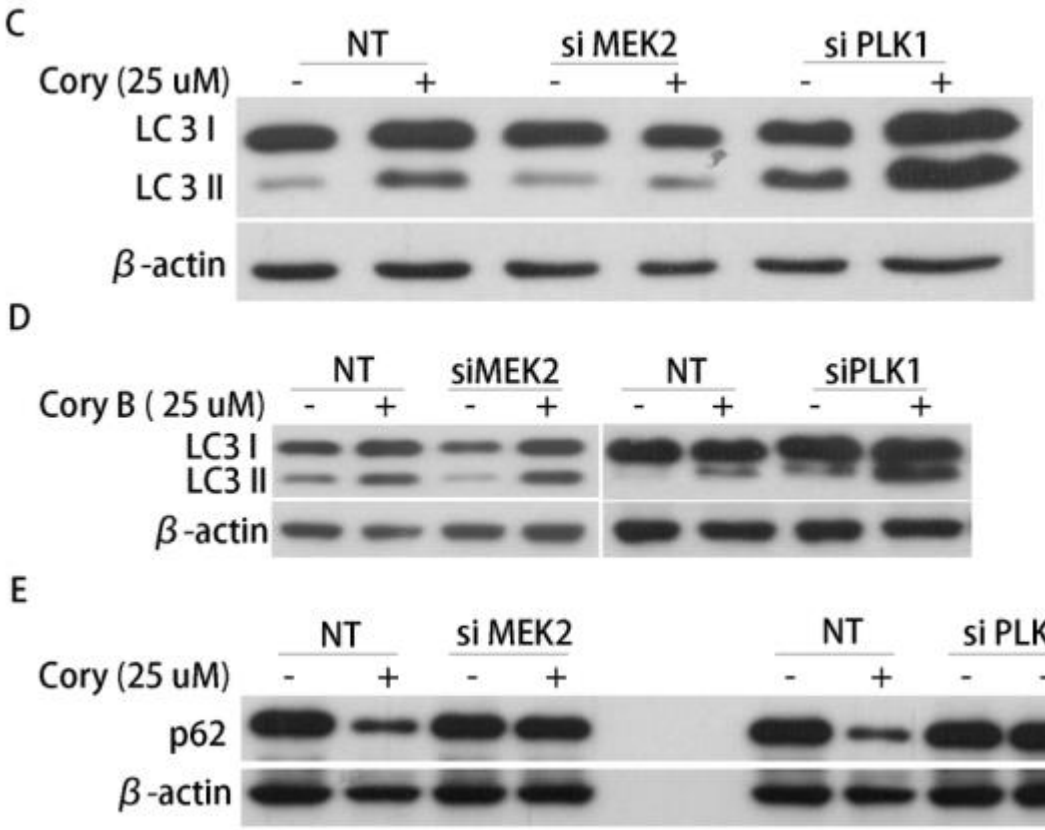






MEK2 & PLK1

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

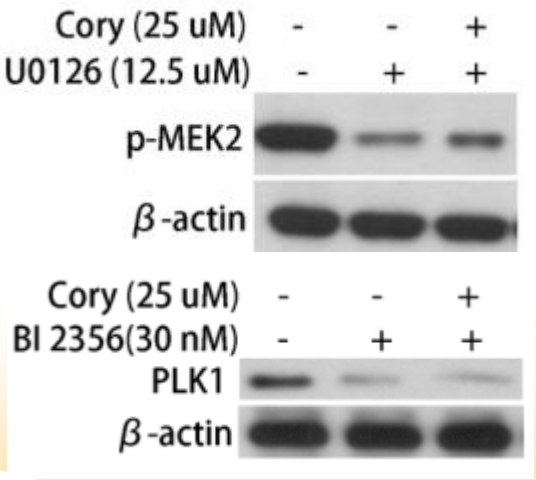


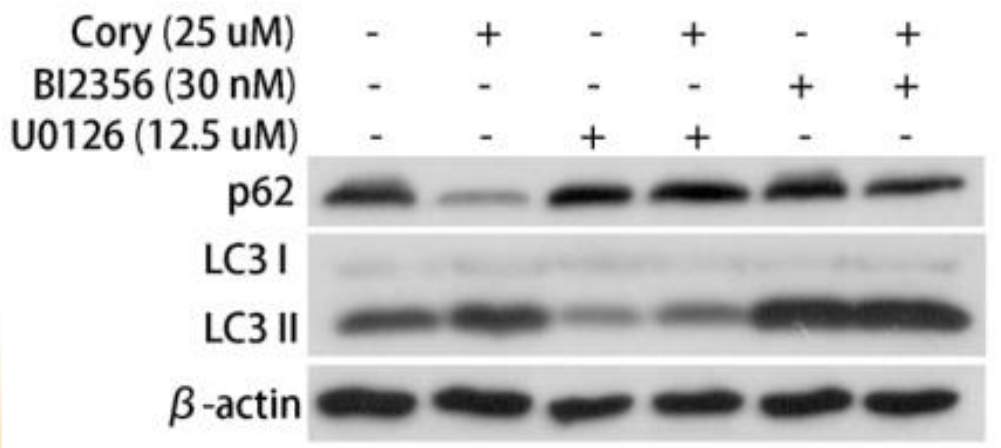
LC3 II ↑: autophagy inhibition & activation

p62↓: autophagic flux ↑

MEK2 & PLK1 activation

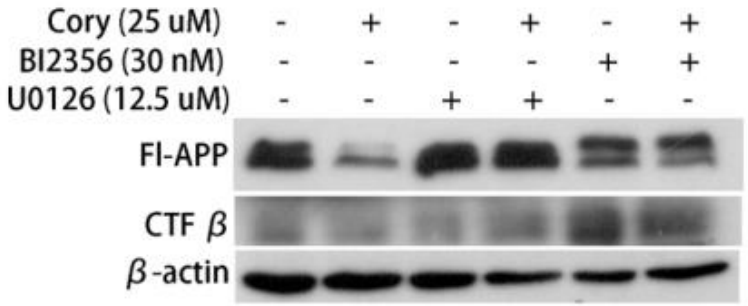
- **Inhibitors: U0126 (MEK2) & BI2356 (PLK1)**
- **The inhibitions of MEK2 and PLK1 both increase p62 and block autophagic flux**

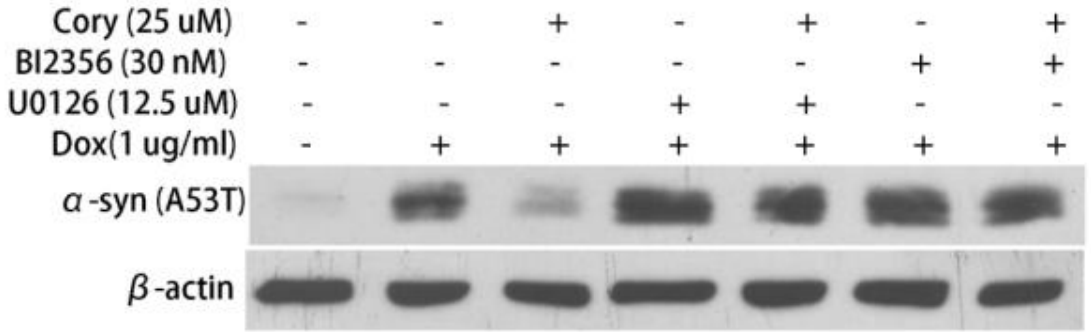


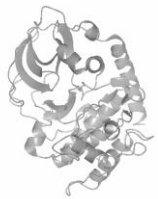


MEK2 & PLK1 in neuronal autophagy

- 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
Leilei Chen

Prof. Xiang-Jiao Yang, McGill

➢ 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?

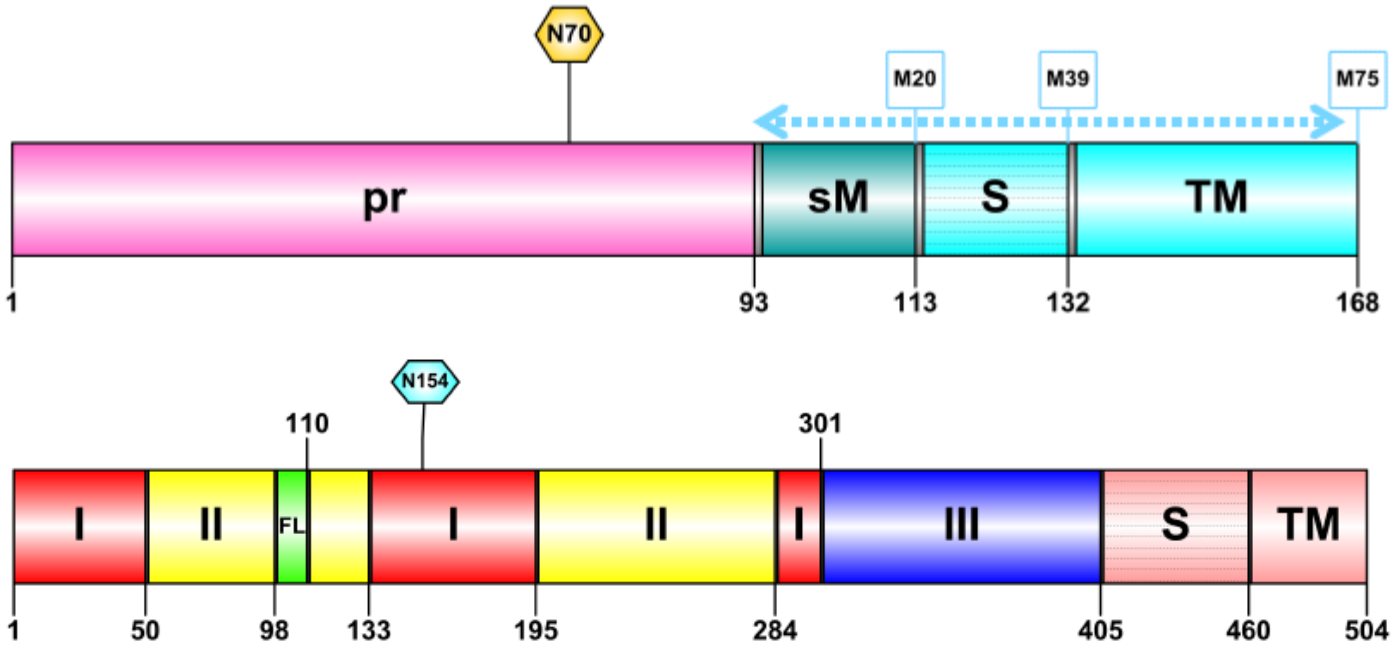
Illustrator for Biological Sequences (IBS)

Science

REPORTS

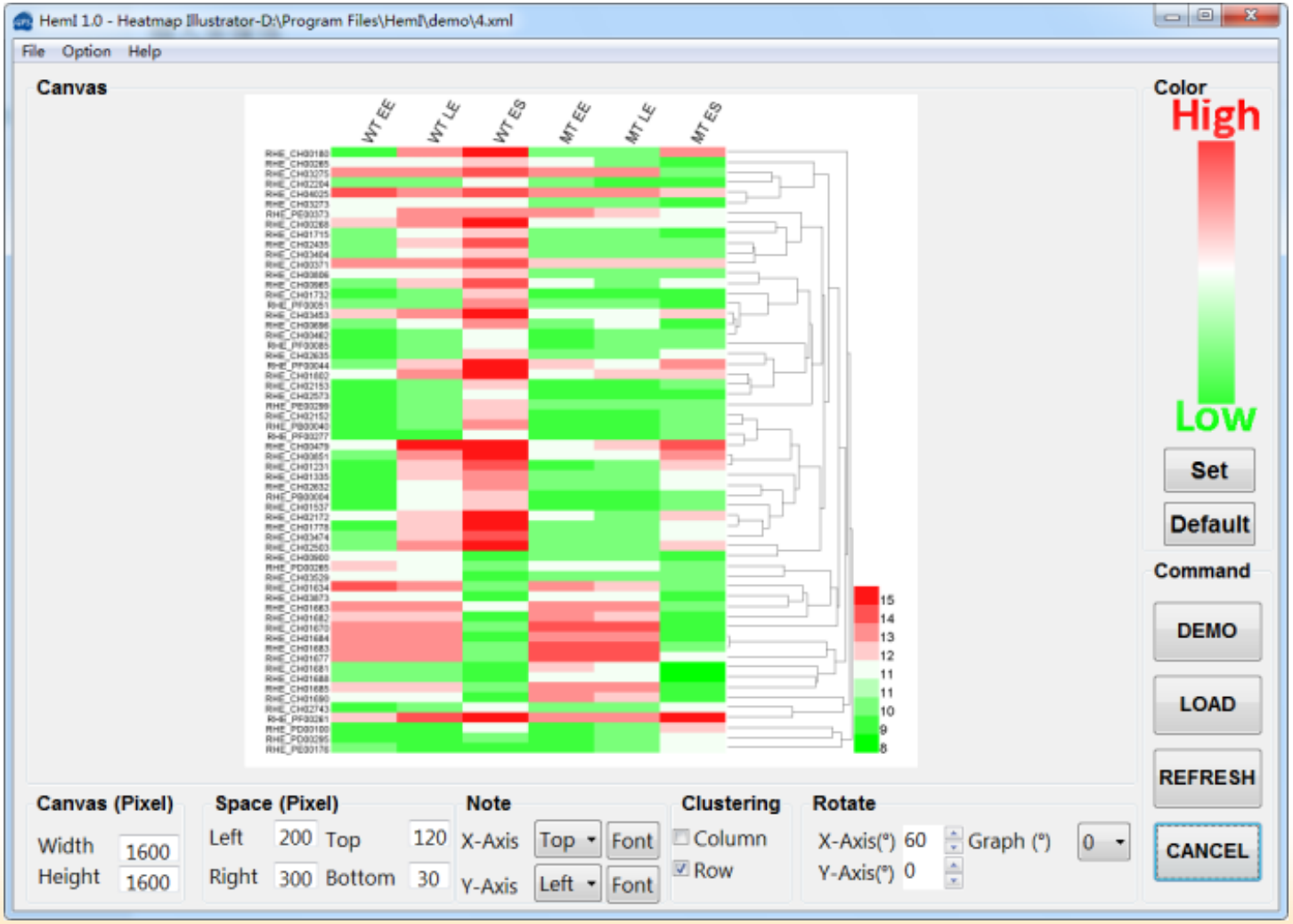
Cite as: Sirohi *et al.*, *Science*
10.1126/science.aaf5316 (2016).

The 3.8 Å resolution cryo-EM structure of Zika virus



➢ **http://ibs.biocuckoo.org**

Heatmap Illustrator (HemI)



➢ **http://hemi.biocuckoo.org**