GPS-PAIL Manual

Prediction of Acetylation on Internal Lysines

Version 2.0.0
10/12/2013

Author: Zexian Liu, Jian Ren & Yu Xue

Contact:
Dr. Zexian Liu, lzx.bioinfo@gmail.com
Dr. Jian Ren, renjian.sysu@gmail.com
Dr. Yu Xue, xueyu@hust.edu.cn
The software is only free for academic research.
The latest version of GPS-PAIL software is available from http://pail.biocuckoo.org
Copyright (c) 2013. The CUCKOO Workgroup. All Rights Reserved.

Index

Statement

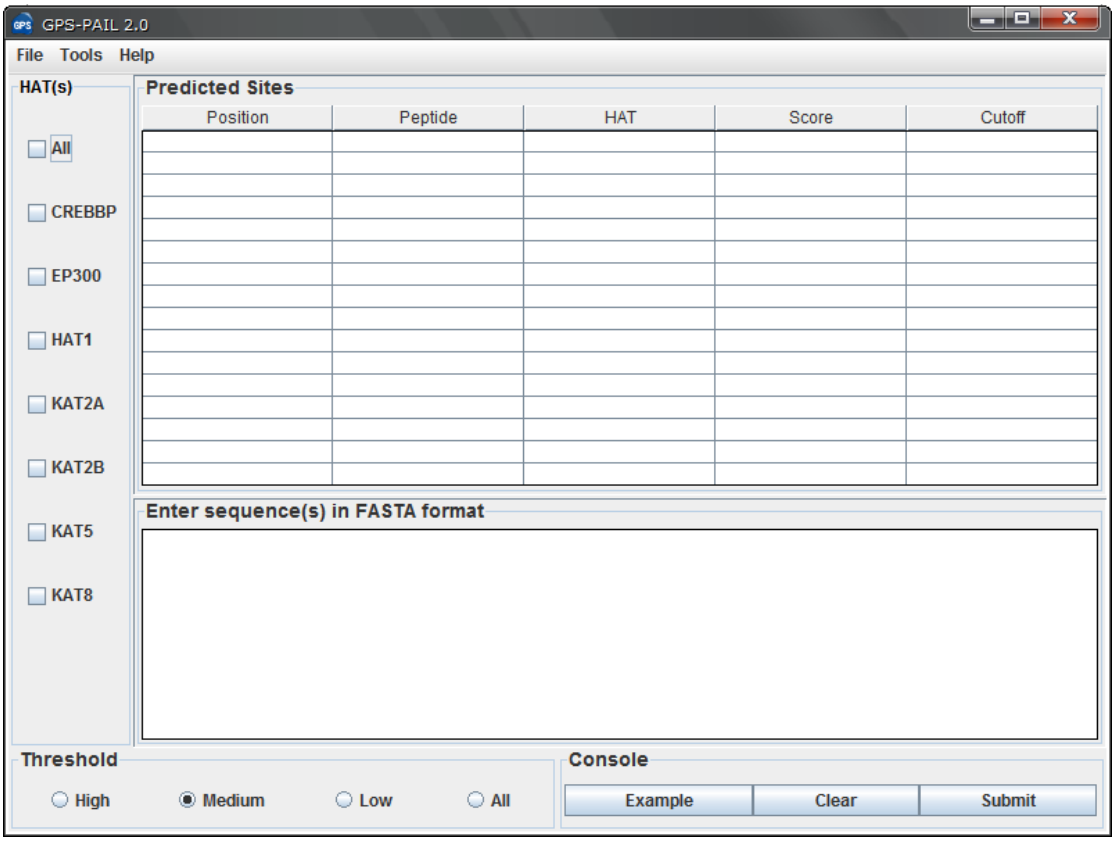
1. **Implementation**. The softwares of the CUCKOO Workgroup are implemented in JAVA (J2SE). Usually, both of online service and local stand-alone packages will be provided.
2. **Availability**. Our softwares are freely available for academic researches. For non-profit users, you can copy, distribute 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 the the National Basic Research Program (973 project) (2012CB910101, and 2013CB933903), Natural Science Foundation of China (31171263 and 81272578), and International Science & Technology Cooperation Program of China (0S2013ZR0003).

Introduction

There are two types of acetylation processes widely occurred in proteins (1-8). The first N^{α}-terminal acetylation is catalyzed a variety of N-terminal acetyltransferases (NATs), which cotranslationally transfer acetyl moieties from acetyl-coenzyme A (Acetyl-CoA) to the α-amino (N^{α}) group of protein amino-terminal residues (1,2). Although N^{α}-terminal acetylation is rare in prokaryotes, it was estimated that about 85% of eukaryotic proteins are N^{α}-terminally modified (1,2). The second type is N^{ε}-lysine acetylation, which specifically modifies ε-amino group of protein lysine residues (3-8). Although N^{ε}-lysine acetylation is less common, it's one of the most important and ubiquitous post-translational modifications conserved in prokaryotes and eukaryotes (1,2). Moreover, the acetylation and deacetylation are dynamically and temporally regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively (4-8).

In 1964, Allfrey *et al*. firstly observed that lysine acetylation of histones plays an essential role in regulation of gene expression (9). Later and recent studies in epigenetics solidified this seminal discovery, and proposed acetylation as a key component of the "histone code" (10,11). Beyond histones, a wide range of non-histone proteins can also be lysine acetylated, and involved in a variety of biological processes, such as transcription regulation (12), DNA replication (13,14), cellular signaling (15,16), stress response (17) and so on. Aberrance of lysine acetylation and deacetylation is associated with various diseases and cancers (5,8,18). In particular, acetylation was demonstrated to be implicated in cellular metabolism and aging (19-21), while one class of NAD^{+} dependent HDACs of sirtuins might be potent drug target for promoting longevity (17,22,23).

In this work, we manuall collected experimentally identified protein lysine acetylation sites for 7 HATs from scientific literature. A previously self-developed GPS (Group-based Prediction System) algorithm was employed with great improvement. We calculated the leave-one-out validation and 4-, 6-, 8-, 10-fold cross-validations to evaluate the prediction performance and system robustness. The online service and stand-alone packages of GPS-PAIL 2.0 were implemented in JAVA and freely available at: http://tsp.biocuckoo.org/.

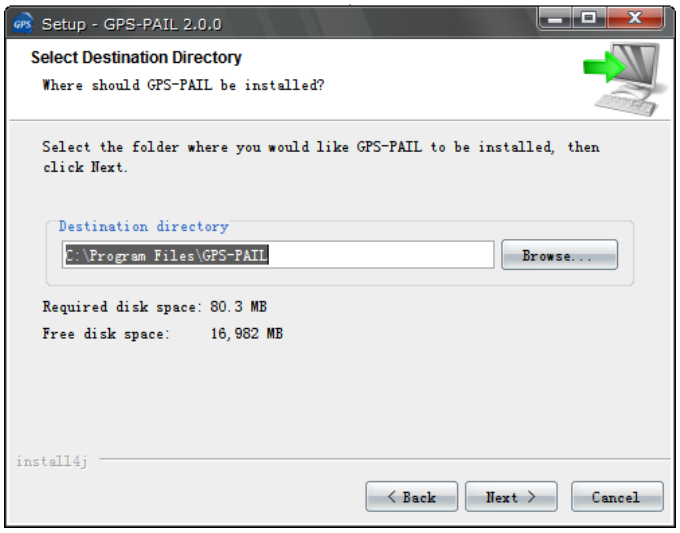


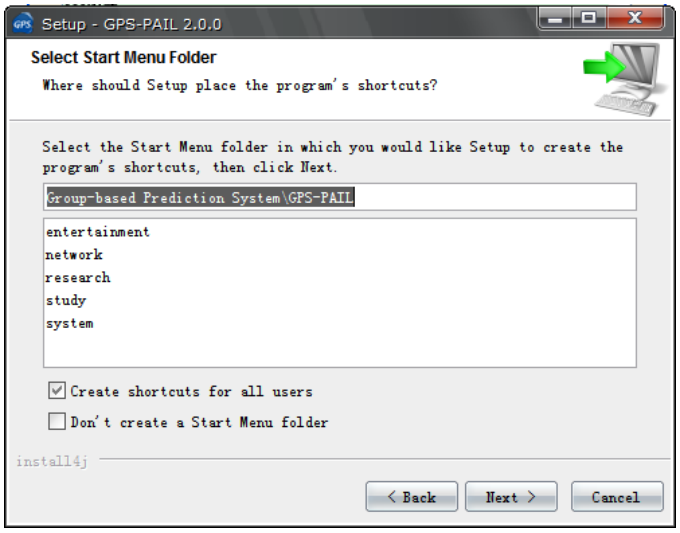
GPS-PAIL 2.0 User Interface

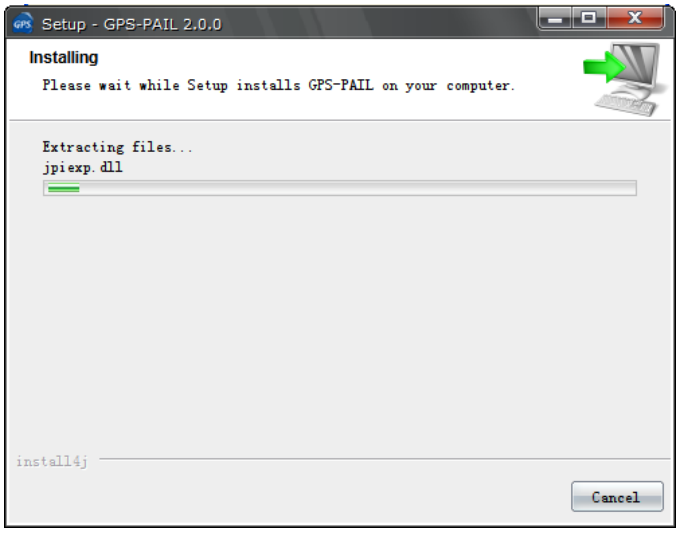
Download & Installation

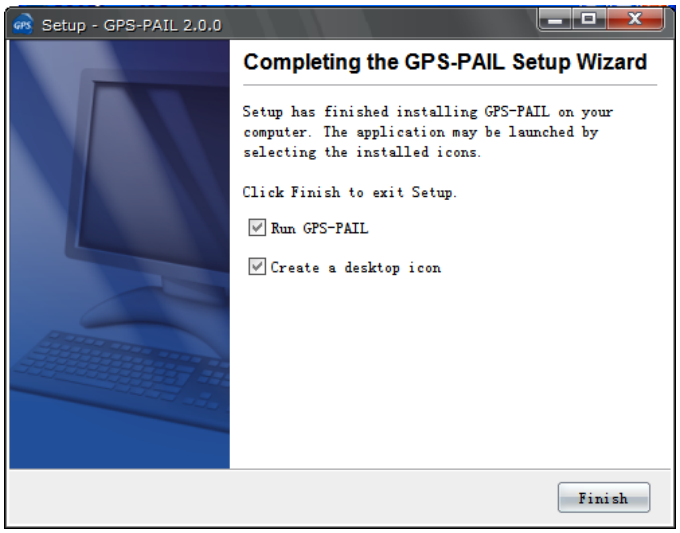
The GPS-PAIL 2.0 was implemented in JAVA (J2SE), and could support three major Operating Systems (OS), including Windows, Linux/Unix or Mac OS X systems. Both of online web service and local stand-alone packages are available from: http://tsp.biocuckoo.org/ . We recommend that users could download the latest release.

Please choose the proper package to download. After downloading, please double-click on the software package to begin installation, following the user prompts through the installation. And snapshots of the setup program for windows are shown below:









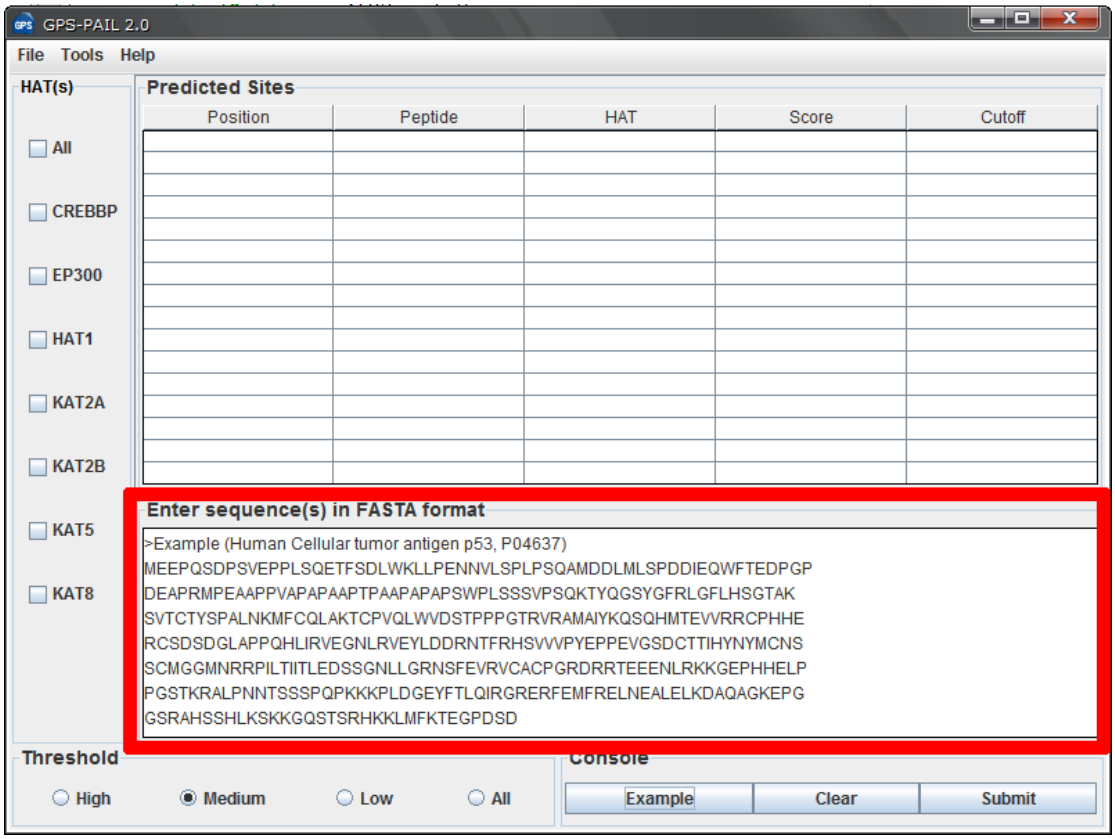
Finally, please click on the **Finish** button to complete the setup program.

Prediction of Acetylation on Internal Lysines

1. A single protein sequence in FASTA format

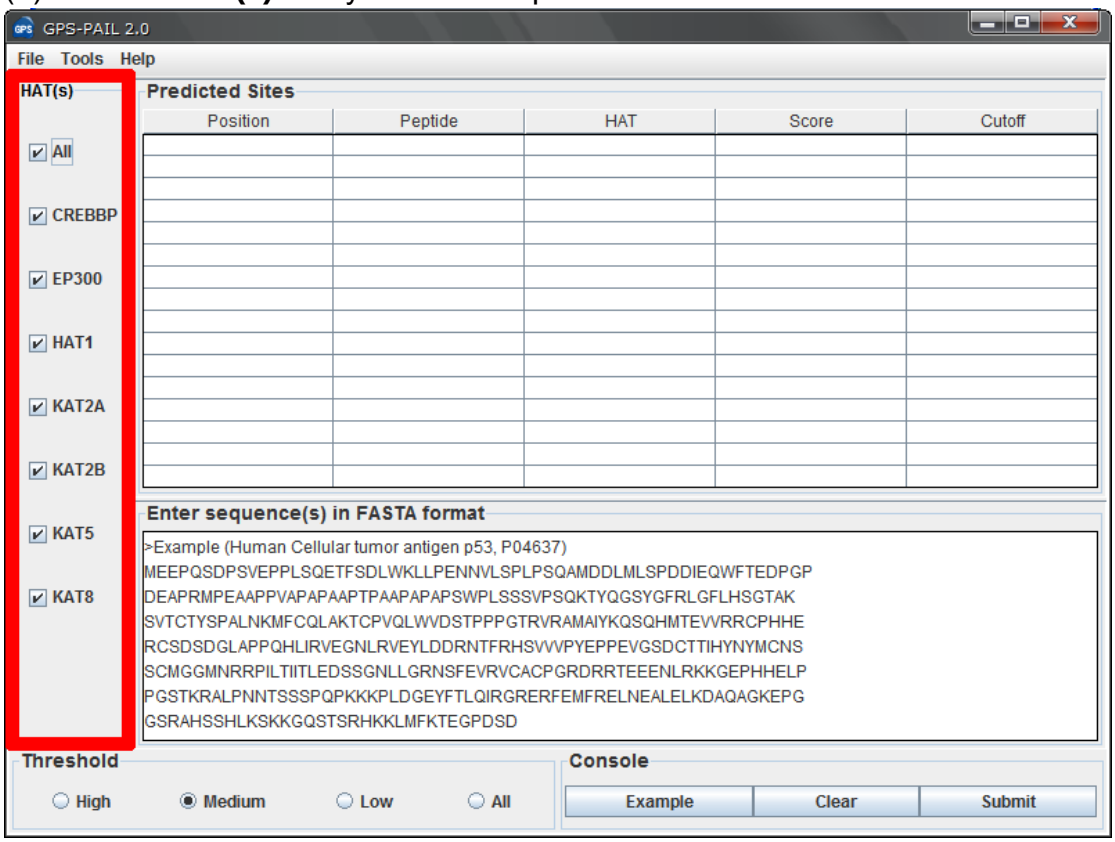
The following steps show you how to use the GPS-PAIL 2.0 to predict lysine acetylation sites for a single protein sequence in FASTA format.

(1) Firstly, please use “Ctrl+C & Ctrl+V” (Windows & Linux/Unix) or “Command+C & Command+V” (Mac) to copy and paste your sequence into the text form of GPS-PAIL 2.0

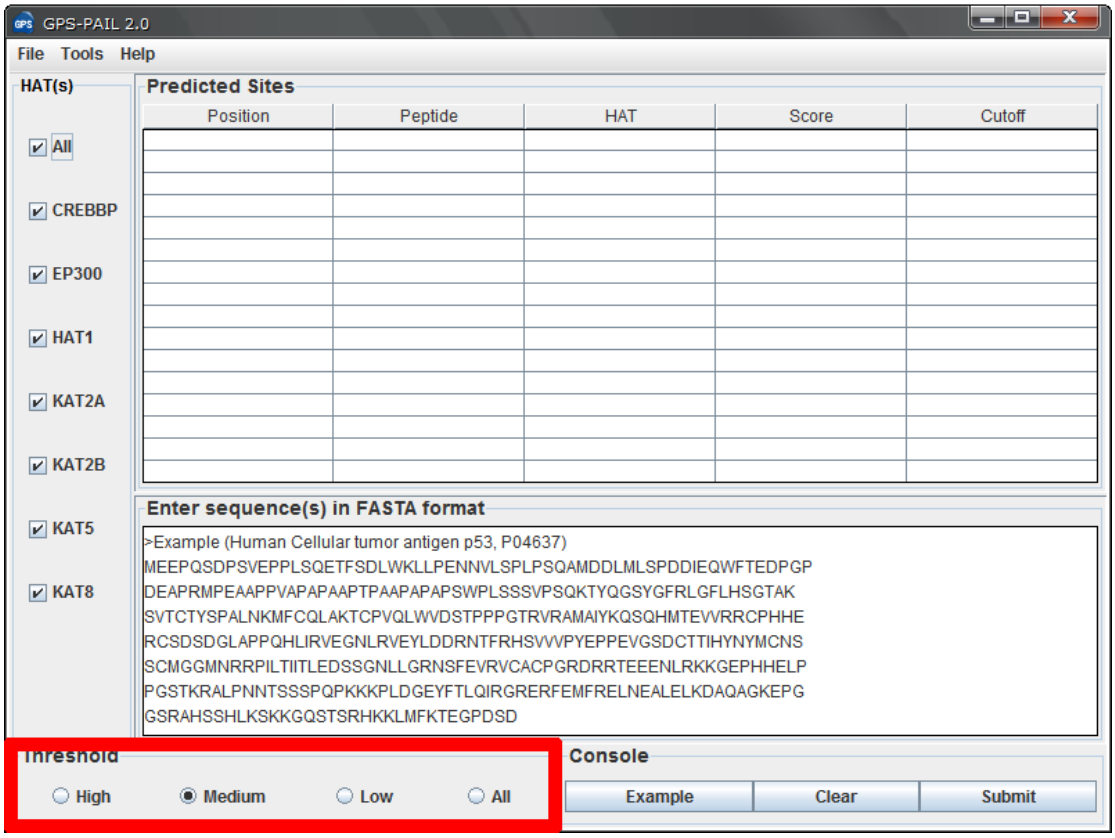


Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

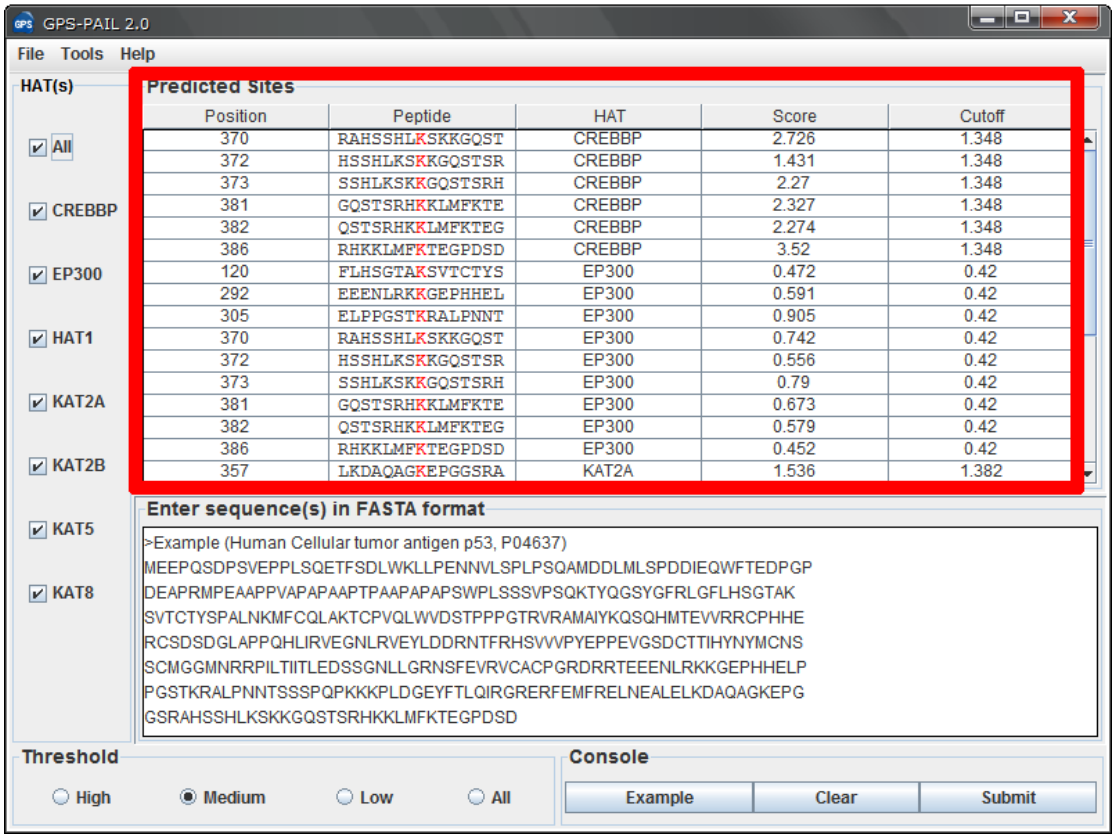
(2) Choose **HAT(s)** that you need for prediction.



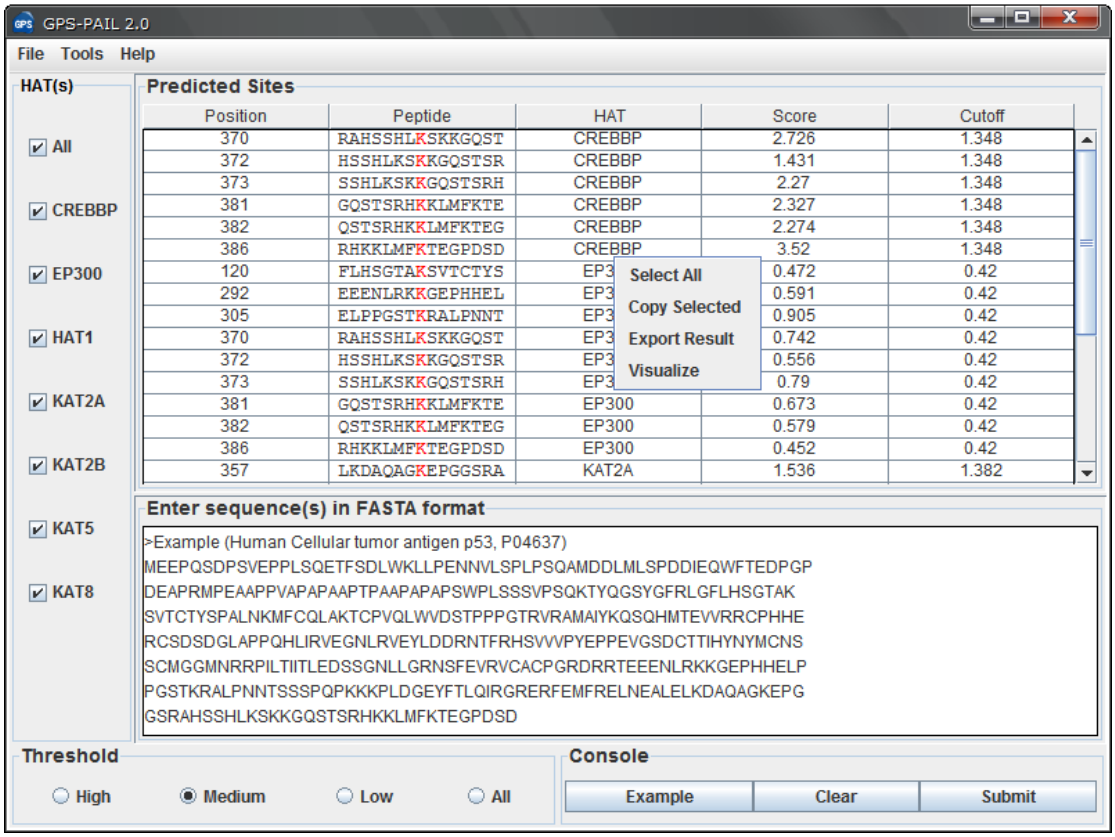
(3) Choose a **Threshold** that you need, the default cut-off is **Medium**.



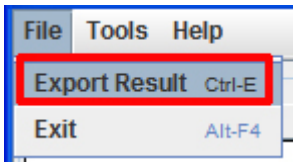
(4) Click on the **Submit** button, then the predicted lysine acetylation sites will be shown.



(5) Then please click on the **RIGHT** button in the prediction 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 Prediction**" to export the prediction results into a tab-delimited text file.



Again, you can also click the "**Export Prediction**" in **File** menu to export the results.

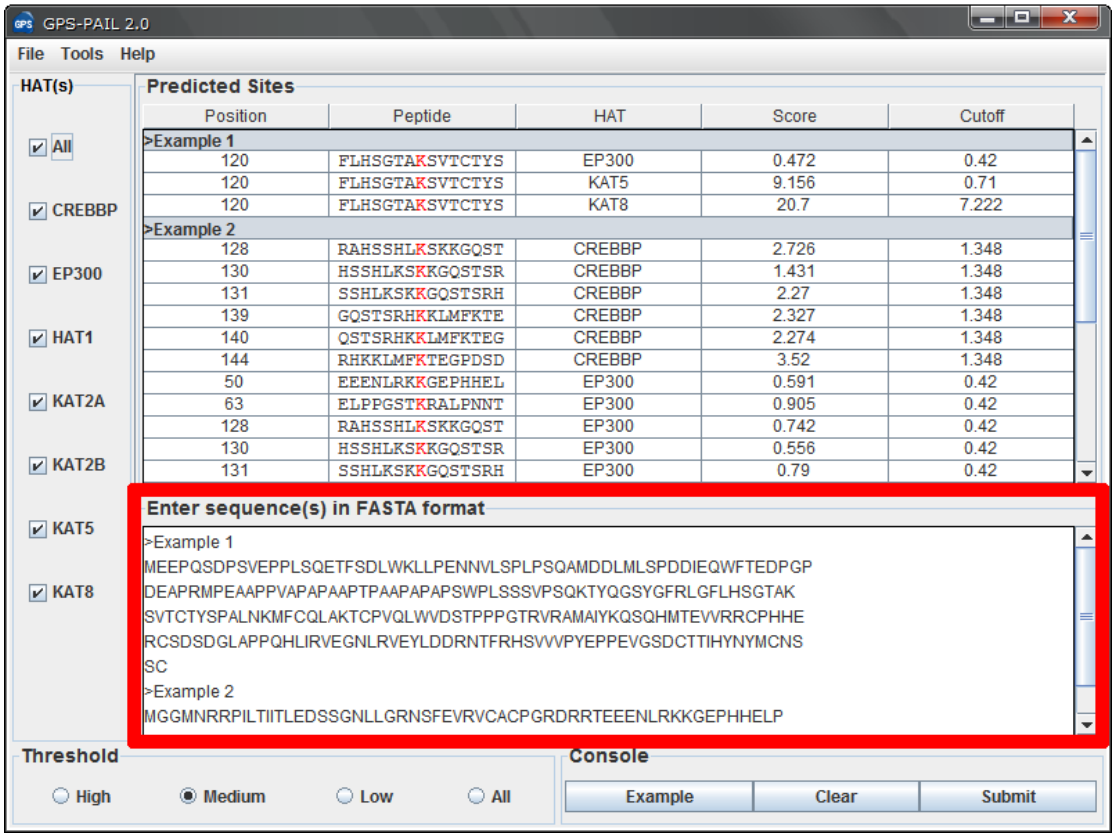


2. Multiple protein sequences in FASTA format

For multiple protein sequences, there are two ways to use the GPS-PAIL 2.0.

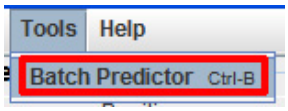
A. Input the sequences into text form directly. (Num. of Seq ≤ 2,000)

If the number of total protein sequences is not greater than 2,000, you can just use "Ctrl+C & Ctrl+V" (Windows & Linux/Unix) or "Command+C & Command+V" (Mac) to copy and paste your sequences into the text form of GPS-PAIL 2.0 for prediction.



B. Use Batch Predictor tool.

If the number of protein sequences is very large, eg., yeast or human proteome, please use the **Batch Predictor**. Please click on the "**Batch Predictor**" button in the **Tools** menu.

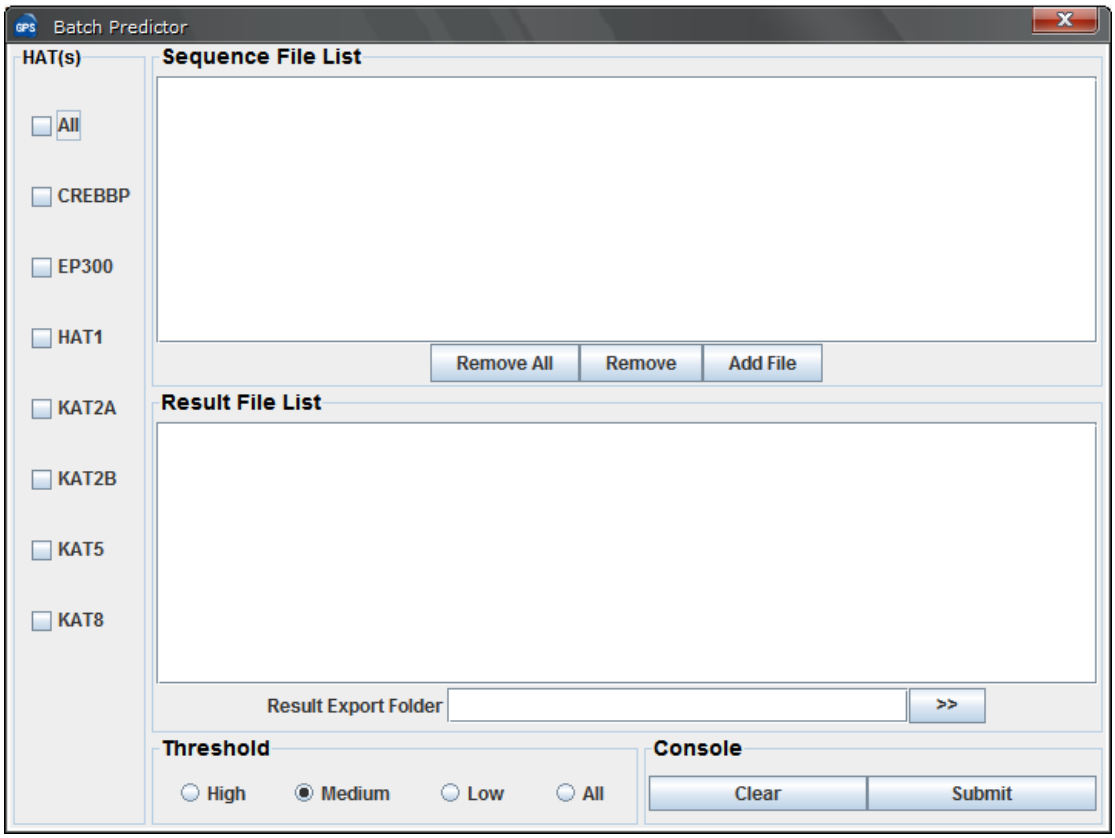


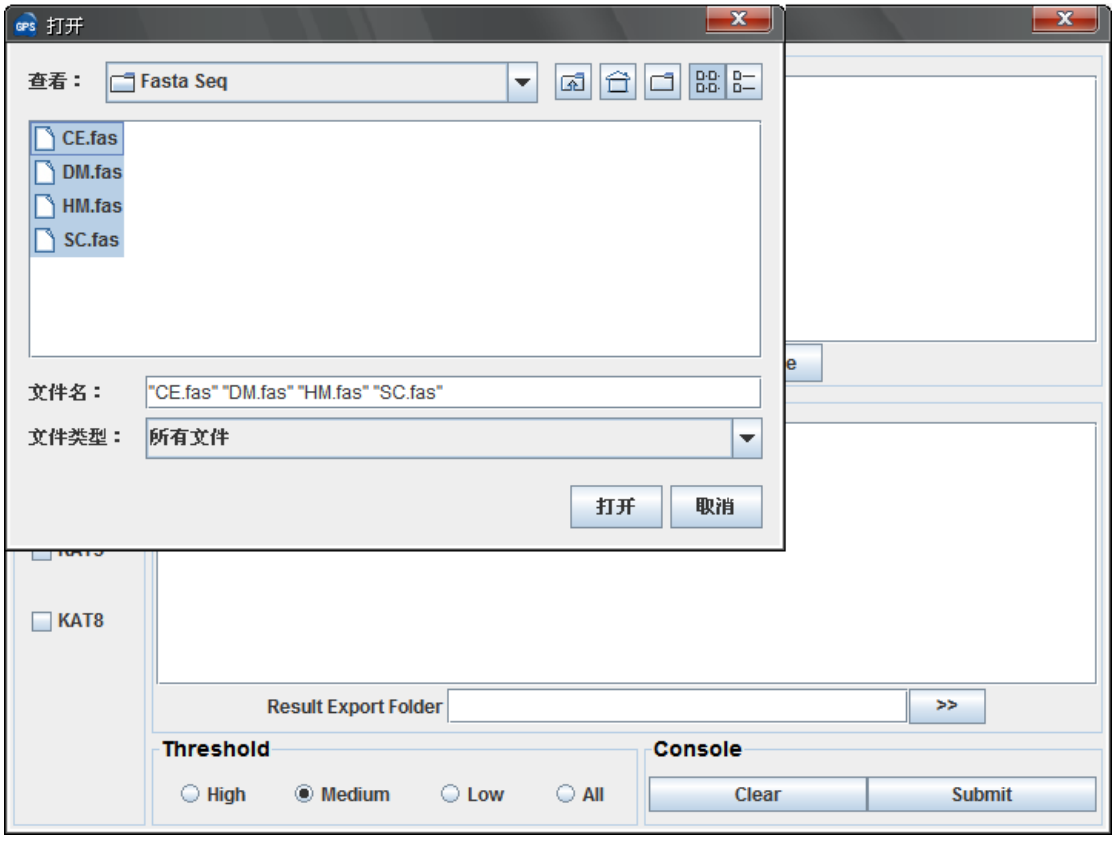
The following steps show you how to use it:

(1) Put protein sequences into one or several files (eg., SC.fas, CE.fas, and etc) with FATSA format as below:

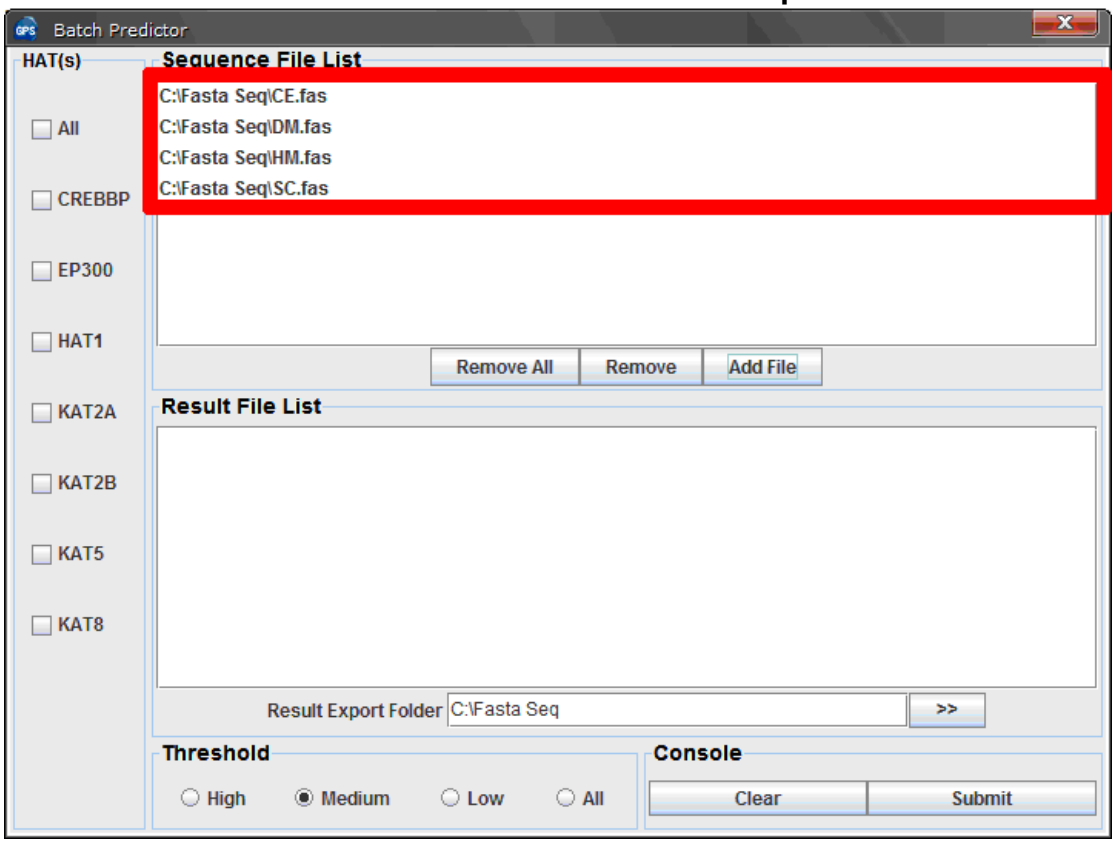
>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXXX
...
Most importantly, the name of each protein should be presented.

(2) Click on the **Batch Predictor** button and then click on the **Add File** button and add one or more protein sequence files in your hard disk.

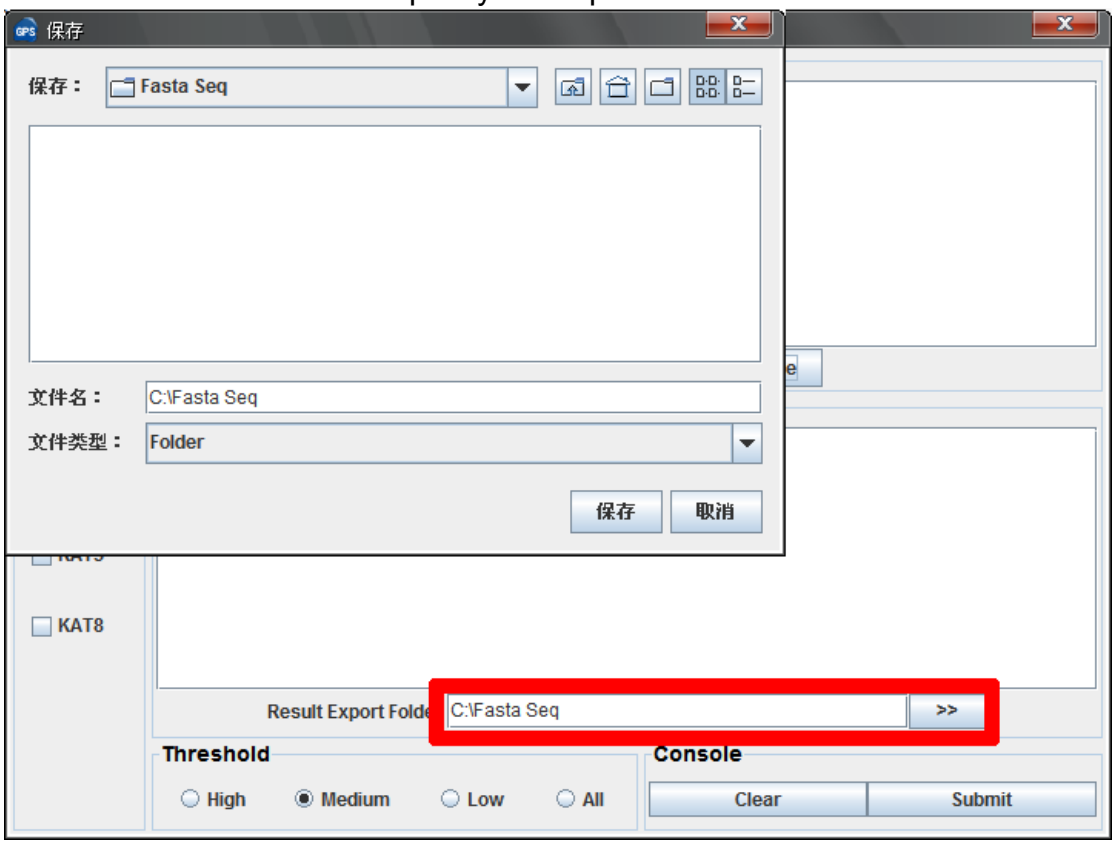




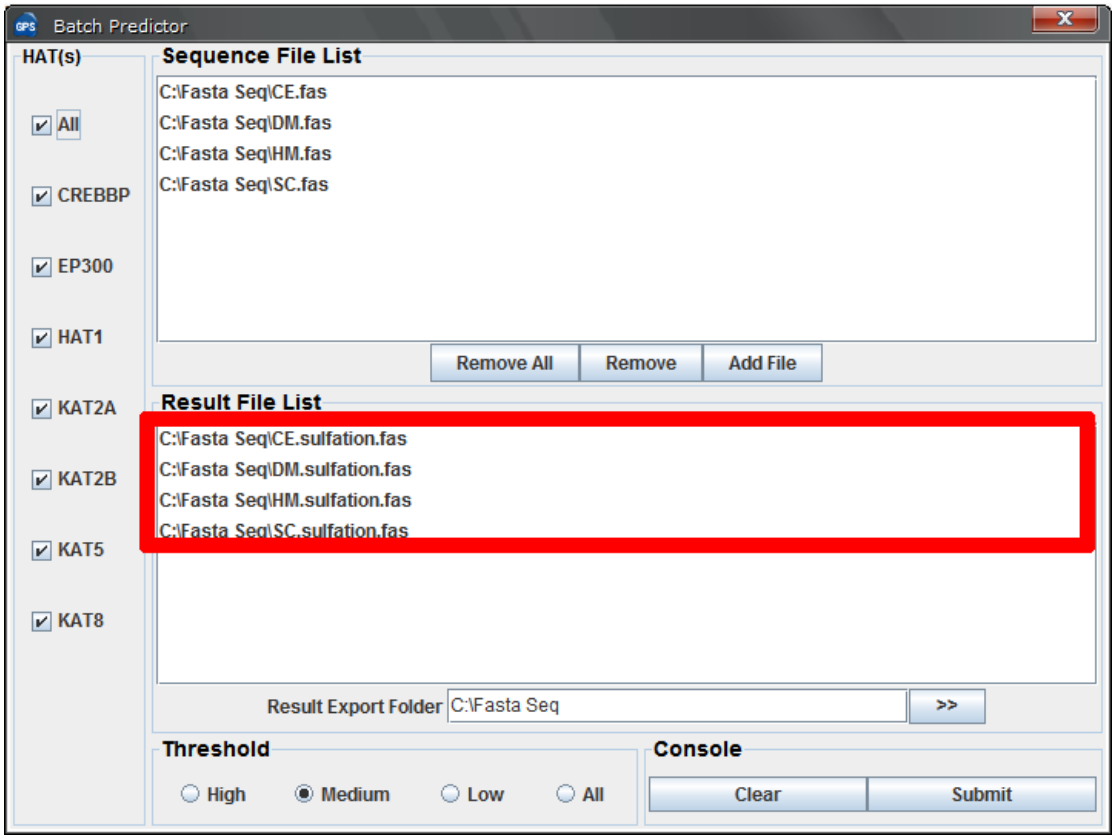
Then the names of added files will be shown in the **Sequence File List**.



(3) The output directory of prediction results should also be defined. Please click on the **>>** button to specify the export fold.



(4) Please choose the HAT(s) and a proper threshold before prediction. Then please click on the **Submit** button, then the **Batch Predictor** begin to process all of the sequence files that have been added to the list. The result of prediction will be export to the **Prediction Export Fold,** and the name of result files will be shown in the **Prediction File List**.



References

1. Polevoda, B. and Sherman, F. (2000) Nalpha -terminal acetylation of eukaryotic proteins. *J Biol Chem*, **275**, 36479-36482.
2. Polevoda, B. and Sherman, F. (2002) The diversity of acetylated proteins. *Genome Biol*, **3**, reviews0006.
3. Smith, K.T. and Workman, J.L. (2009) Introducing the acetylome. *Nat Biotechnol*, **27**, 917-919.
4. Yang, X.J. and Seto, E. (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol*, **9**, 206-218.
5. Yang, X.J. and Seto, E. (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, **26**, 5310-5318.
6. Shahbazian, M.D. and Grunstein, M. (2007) Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem*, **76**, 75-100.
7. Lee, K.K. and Workman, J.L. (2007) Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol*, **8**, 284-295.
8. Yang, X.J. (2004) The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res*, **32**, 959-976.
9. Allfrey, V.G., Faulkner, R. and Mirsky, A.E. (1964) Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A*, **51**, 786-794.
10. Yang, X.J. and Seto, E. (2008) Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol Cell*, **31**, 449-461.
11. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.
12. Yuan, Z.L., Guan, Y.J., Chatterjee, D. and Chin, Y.E. (2005) Stat3 dimerization regulated by reversible acetylation of a single lysine residue. *Science*, **307**, 269-273.
13. Terret, M.E., Sherwood, R., Rahman, S., Qin, J. and Jallepalli, P.V. (2009) Cohesin acetylation speeds the replication fork. *Nature*, **462**, 231-234.
14. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V. and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, **325**, 834-840.
15. Walkinshaw, D.R., Tahmasebi, S., Bertos, N.R. and Yang, X.J. (2008) Histone deacetylases as transducers and targets of nuclear signaling. *J Cell Biochem*, **104**, 1541-1552.
16. Spange, S., Wagner, T., Heinzel, T. and Kramer, O.H. (2009) Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int J Biochem Cell Biol*, **41**, 185-198.
17. Brunet, A., Sweeney, L.B., Sturgill, J.F., Chua, K.F., Greer, P.L., Lin, Y., Tran, H., Ross, S.E., Mostoslavsky, R., Cohen, H.Y. *et al.* (2004) Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science*, **303**, 2011-2015.
18. Kim, S.C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L. *et al.* (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell*, **23**, 607-618.
19. Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H., Xie, L., Zhao, W., Yao, Y. *et*

al. (2010) Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science*, **327**, 1004-1007.

20. Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J., Zhou, L., Zeng, Y., Li, H. *et al.* (2010) Regulation of cellular metabolism by protein lysine acetylation. *Science*, **327**, 1000-1004.
21. Nakamura, A., Kawakami, K., Kametani, F., Nakamoto, H. and Goto, S. (2010) Biological significance of protein modifications in aging and calorie restriction. *Ann N Y Acad Sci*, **1197**, 33-39.
22. Wang, J., Fivecoat, H., Ho, L., Pan, Y., Ling, E. and Pasinetti, G.M. (2010) The role of Sirt1: at the crossroad between promotion of longevity and protection against Alzheimer's disease neuropathology. *Biochim Biophys Acta*, **1804**, 1690-1694.
23. Cohen, H.Y., Miller, C., Bitterman, K.J., Wall, N.R., Hekking, B., Kessler, B., Howitz, K.T., Gorospe, M., de Cabo, R. and Sinclair, D.A. (2004) Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. *Science*, **305**, 390-392.

Release Note

1. April 12, 2012, the online service and the local stand-alone packages of GPS-PAIL 2.0 were released.