

WERAM: a database of writers, erasers and readers of histone acetylation and methylation in eukaryotes

Yang Xu[†], Shuang Zhang[†], Shaofeng Lin, Yaping Guo, Wankun Deng, Ying Zhang and Yu Xue^{*}

Key Laboratory of Molecular Biophysics of Ministry of Education, College of Life Science and Technology and the Collaborative Innovation Center for Brain Science, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Received August 06, 2016; Revised September 28, 2016; Editorial Decision October 16, 2016; Accepted October 18, 2016

ABSTRACT

In this work, we developed a database WERAM (<http://weram.biocuckoo.org/>) for histone acetyltransferases, histone deacetylases, histone methyltransferases, histone demethylases and acetyl- or methyl-binding proteins, which catalyze, remove and recognize histone acetylation and methylation sites as ‘writers’, ‘erasers’ and ‘readers’, and synergistically determine the ‘histone code’. From the scientific literature, we totally collected over 580 experimentally identified histone regulators from eight model organisms, including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*. We also collected ~900 site-specific regulator-histone relations from the eight species. According to the experimental evidence, known histone regulators were classified into distinct families. To computationally detect more proteins in eukaryotes, we constructed hidden Markov model (HMM) profiles for histone regulator families. For families without HMM profiles, we also conducted orthologous searches. Totally, WERAM database contained more than 20 thousand non-redundant histone regulators from 148 eukaryotes. The detailed annotations and classification information of histone regulators were provided, together with site-specific histone substrates if available.

INTRODUCTION

As an important epigenetic mechanism, post-translational modifications (PTMs) of histones or histone modifications combinatorially and dynamically constitute ‘histone code’, which accurately regulates a variety of DNA-templated pro-

cesses such as gene transcription, chromosome segregation and DNA repair (1–8). Acetylation and methylation are two predominant types of histone modifications. Histone acetylation neutralizes the positive charges of lysine residues, disrupts the association between histones and DNA, and is a hallmark of open chromatin that can be more accessible to DNA and RNA polymerases as well as transcription factors, resulting in the activation of gene transcription (3,8–11). Histone methylation mainly occurs on arginine and lysine residues, enhances the basicity and hydrophobicity of histone tails, influences the interaction affinity between DNA and transcription factors, and can either activate or repress gene expression (3,8,11–14). For example, the methylation of H3K4 and H3K36 is closely associated with transcriptional activation, whereas methylated H3K9 and H3K27 are highly correlated with gene repression (11,14).

Histone acetylation and methylation are catalyzed, cleared and recognized by histone modifying enzymes (‘writers’), histone demodifying enzymes (‘erasers’), and acetyl- or methyl-binding proteins (‘readers’), respectively (3,15–17). For example, histone acetyltransferases (HATs), histone deacetylases (HDACs) and bromodomains can deposit, remove or interact with specific acetylation sites (2,6,7,9,10,18,19). For methylation, histone methyltransferases (HMTs), histone demethylases (HDMs) and a variety of functional domains such as chromodomain, PHD, ZF-CW and Tudor domains act as writers, erasers and reader proteins of methylation sites (6,7,9,12,18,20). Histone acetylation and methylation are critically and precisely regulated, whereas the aberrances or mutations of histone modifications or regulators are highly associated with memory impairment, developmental defects and cancers (11,13–16,21). In this regard, the identification of histone modifications, regulators and site-specific regulator-histone relations is fundamental for understanding the molecular mechanisms and regulatory roles of the histone code.

^{*}To whom correspondence should be addressed. Tel: +86 27 8779 3903; Fax: +86 27 8779 3172; Email: xueyu@hust.edu.cn

[†]These authors contributed equally to the paper as first authors.

Recently, due to the rapid progress in the development of high-throughput techniques such as quantitative mass spectrometry and highly potent antibodies (22,23), a large number of new histone PTMs including acetylation and methylation sites were characterized (1,4,20,24). Thus, the collection of histone modifications together with their cognate regulators has emerged to be a great challenge. Currently, there have been four major databases developed for histones and histone modifications, including HHMD, Hlstone, db-HiMo and HistoneDB (25–28). For example, HHMD collected 43 location-specific histone modifications in humans, focusing on the storage and integration of histone modification datasets (25), while Hlstone contained 55 human histones, 106 site-specific PTMs and 152 histone-modifying enzymes (26). The dbHiMo mainly focused on the identification of histone-modifying enzymes in fungi, whereas readers and site-specific histone modifications were not included (27). HistoneDB maintained histone sequences, structural annotations, for the exploration of histones and their variants (28). These data resources mainly focused on the collection of histone sequences and structures, or histone modifications in human and fungi, whereas an integrative database in eukaryotes is still lacking.

From the scientific literature, we first collected 584 experimentally identified histone regulators, including 72 HATs, 97 HDACs, 116 acetyl-readers, 112 HMTs, 76 HDMs and 156 methyl-readers from eight model organisms including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* (Supplementary Table S1). For a better understanding of the histone code, we also collected 904 known site-specific regulator-histone relations of the eight species. Based on previously established rationales (2,6–10,12,18–20,29), we classified all known HATs, HDACs, acetyl-readers, HMTs, HDMs and methyl-readers into distinct protein families. In our results, there were 15 families for the histone acetylation system, and 32 families for histone methylation regulators. Using HMMER (30), we totally constructed 13 and 30 hidden Markov models (HMMs) for histone acetylation and methylation regulators at the family level, respectively. Then we systematically characterized 1077 HATs, 2408 HDACs, 3901 acetyl-readers, 3990 HMTs, 1610 HDMs and 10 685 methyl-readers from 148 eukaryotic species using the HMM profile of each family. Moreover, for four families without HMMs, the ortholog search was conducted to identify potential histone regulators. The detailed annotations from the Ensembl (31) and UniProt (32) databases were integrated, and the classification information was also provided. Finally, an integrative database of writers, erasers and readers of acetylation and methylation in eukaryotes (WERAM) was developed with 20 033 non-redundant histone regulators, including 1337 HATs, 2504 HDACs, 3901 acetyl-readers, 4409 HMTs, 1610 HDMs and 10 949 methyl-readers.

CONSTRUCTION AND CONTENT

Data collection

Because too many scientific papers were published on histone modifications, here we mainly focused on the cu-

ration of the literature published since 2011. For histone acetylation, we used the PubMed Advanced Search Builder and built three filters, including ‘(histone acetyltransferase) NOT (histone acetyltransferase inhibitor)’, ‘(((histone deacetylase [Title/Abstract]) OR HDAC [Title/Abstract]) NOT (histone deacetylase inhibitors) NOT histone deacetylase inhibitor) NOT histone deacetylase inhibition)’, and (histone acetylation) AND (bromodomain or tandem PHD), and retrieved approximately 2500, 2000 and 150 papers, respectively. For histone methylation, we directly searched the PubMed using multiple keywords, such as ‘histone methyltransferase’, ‘histone demethylase’, ‘histone methylation reader’, ‘histone methylation’ and ‘histone modification’, and retrieved over 3000 abstracts in total. In addition, we also considered several important review articles published before 2011 (2,8,11,12,17,19,29). From the abstracts or full-length manuscripts, we manually curated 248 and 336 experimentally identified acetylation and methylation regulators, respectively (Supplementary Table S1).

For the 584 known histone regulators, the information of protein domains was taken from the annotations in UniProt database (32). The functional domains of all writers, erasers and readers were further examined by searching the Pfam database (33). Furthermore, we downloaded the complete proteome sets for 148 eukaryotes including 68 animals, 39 plants and 41 fungi, from Ensembl (31) (release version 84, <http://www.ensembl.org/>), under the directory of ‘/pub/release-84/fasta’, EnsemblPlants (release version 31, <http://plants.ensembl.org/>) and EnsemblFungi (release version 31, <http://fungi.ensembl.org/>), respectively. Because multiple variant nucleotide sequences or peptides can be originated from a single gene, we further used Ensembl Gene ID as the unique accession to eliminate the redundancy. For multiple alternatively splicing isoforms of a single gene, only the longest one was reserved.

The classification of histone regulators

An accurate classification map of known histone regulators is greatly helpful for further identification of new proteins. As previously described (2,6–10,18,19,29), we manually classified collected HATs, HDACs and acetyl-readers into 15 families: (i) HATs have eight families, including p300_CBP, ELP3, GCN5, HAT1, HPA2, GNAT_other, MYST and HAT_other (unclassified HATs); (ii) HDACs contain five families, including Class-I, Class-II, SIR2, Class-IV and HD2; (iii) The acetyl-readers (Ac.Reader) were classified into two families such as Bromodomain, and Tandem-PHD containing at least two PHD domains in proteins (Figure 1).

Also, based on experimental evidence (6–9,12,18,20), we manually classified known HMTs, HDMs and methyl-reader into 32 families: (i) HMTs were clustered into 10 families, including SET1, SET2, RIZ, EZ, SUV39, SUV4-20, SMYD, PRMT, nonSET and HMT_other (unclassified HMTs); (ii) HDMs have eight families, including JARID, JHDM1, JHDM2, JHDM3.JMJD2, PHF2.PHF8, UTX.UTY, JmjC_only and LSD1.KDM1. The JmjC_only family contains HDMs with JmjC domains but not to be classified into any other families

HAT		HMT		Me_Reader	
p300_CBP	525.8	SET1	92.0	ADD	182.7
ELP3	299.7	SET2	144.7	Ankyrin	487.2
GCN5	164.9	RIZ	112.0	BAH	133.5
GNAT_other	213.1	EZ	193.4	Chromodomain	64.8
HAT1	178.3	SUV39	97.6	PHD	32.4
HPA2	226.1	SUV4-20	201.2	PWWP	56.7
MYST	318.4	SMYD	392.1	TTD	148.4
HDAC		PRMT	371.3	Tudor	15.2
Class-I	23.2	nonSET	487.1	WD40	300.4
Class-II	306	HDM		ZF-CW	64.6
Class-IV	487.8	JARID	304.9	Chromo-Barrel	111.1
SIR2	200.6	JHDM1	240.5	MBT	142.8
Ac_Reader		JHDM2	146.3	DCD	64.6
Bromodomain	22.7	JHDM3_JMJD2	206.3		
Tandem-PHD	107.7	JmjC_only	128.2		
Acetylation Families		LSD1_KDM1	704.3		
		PHF2_PHF8	260.4		
		UTX_UTY	311.3		
		Methylation Families			

Figure 1. The detailed classifications of HATs, HDACs, Ac.Readers, HMTs, HDMs and Me.readers, together with a cut-off value of the proteome-wide identification for each family. The detailed parameters used for the hmmsearch program together with a summary of the testing data set were shown in Supplementary Table S2.

of the JmjC superfamily; (iii) The methyl-reader proteins (Me_Reader) were grouped into 14 families, including ADD, Ankyrin, BAH, Chromodomain, PHD, PWWP, TTD, Tudor, WD40, ZF-CW, Chromo-Barrel, MBT, DCD and Me_Reader_other (unclassified methyl-readers) (Figure 1).

Genome-wide identification of HATs, HDACs, HMTs, HDMs and reader proteins

For families with ≥3 member genes, the catalytic domain sequences of HATs, HDACs, HMTs and HDMs, or PTM-binding domain sequences of acetyl-readers and methyl-readers were first aligned for each family separately with MUSCLE (<http://www.drive5.com/muscle/>, version 3.8.31), an extensively used tool for multiple sequence alignment (34). An illustration of aligned domain sequences was shown for four histone regulator families, including HAT/GCN5, HDAC/Class-IV, HMT/SET2 and HDM/JHDM3_JMJD2 (Supplementary Figure S1). HMM models of 569 known histone regulators were then constructed at the family level using the hmmbuild program of the HMMER 3.1b2 package (<http://hmmerr.org/>) (30). We totally constructed 13 and 30 HMM profiles for acetylation and methylation families, respectively.

For the genome-wide identification of histone regulators, the hmmsearch program of HMMER 3.1b2 was used (30). The hmmsearch calculates both the log-odds likelihood scores and E-values for given protein sequences. Because E-values depend on the database size and generate inconsistent results when the database is changed, here realistic constant log-odds likelihood scores were adopted as the threshold values. To evaluate the prediction performance of the HMM identification and specify the cut-off value for each family, the 569 collected histone regulators used for constructing HMM models were taken as the testing dataset. For each family, the known annotated proteins were regarded as positive data (P), while all proteins in other families were taken as negative data (N) (Supplementary Ta-

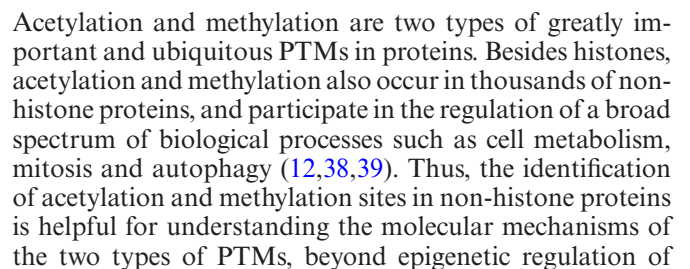
ble S2). Two measurements of sensitivity (Sn) and specificity (Sp) were defined and calculated as shown below:

$$Sn = \frac{TP}{TP + FN}, \text{ and } Sp = \frac{TN}{TN + FP}$$

First, the self-consistency validation was performed directly with the positive data and negative data for each family. To further evaluate the prediction robustness, the leave-one-out validation was also carried out. As an example, the receiver operating characteristic (ROC) curves were drawn and area under ROC values were calculated for eight families of acetylation regulators (Supplementary Figure S2A), and eight families of methylation regulators (Supplementary Figure S2B), respectively. Our results demonstrated that HMM predictions are both accurate and robust (Supplementary Figure S2). To promise that all known histone regulators can be correctly identified and classified in leave-one-out validations (Sn = 100%), we manually selected a log-odds likelihood score as the threshold for each family (Figure 1). The statistics of the testing data set together with the detailed parameters used for hmmsearch were shown (Supplementary Table S2). Then we used hmmsearch to search potential histone regulators in 148 eukaryotes. Because multiple variant peptides can be generated from a single gene, here we used the Ensembl Gene ID as the unique accession to avoid any redundancy. For a given gene, only the protein with the highest log-odds likelihood score was retained as the representative sequence.

For the four families without HMM profiles (Other of HATs, HD2 of HDACs, Other of HMTs and Other of methyl-readers), we further performed orthologous searches to identify 260 HATs, 96 HDACs, 419 HMTs and 264 methyl-readers from 148 eukaryotic organisms, respectively. As previously described (35), the strategy of Reciprocal Best Hits (RBH) was chosen, and the blastall program in the BLAST package was utilized (36). All protein sequences and HMM profiles of histone regulators can be freely downloaded at: <http://weram.biocuckoo.org/download.php>.

In total, we identified and integrated 20 033 histone regulators, including 1337 HATs, 2504 HDACs, 3901 acetyl-readers, 4409 HMTs, 1610 HDMs, 10 949 methyl-readers from 148 eukaryotic species (Supplementary Table S3), whereas a heatmap of the results was visualized using the HemI program (<http://hemi.biocuckoo.org/>) (Supplementary Figure S3) (37). Averagely, there were 90.4 PTM enzymes and 161.0 readers per animal, 73.4 enzymes and 75.9 readers per plant, and 20.7 enzymes and 23.0 readers per fungus (Supplementary Figure S3). Although the numbers of PTM enzymes and readers are quite similar in plants and fungi, animals generally contain more reader proteins than writers and erasers. We also compared WERAM to an existing database dbHiMo, which mainly focused on the identification of histone-modifying enzymes but not reader proteins in fungi (27). For eight eukaryotic organisms, we totally identified 1260 histone regulators, whereas dbHiMo only contained 168 histone-modifying enzymes (Supplementary Table S4).



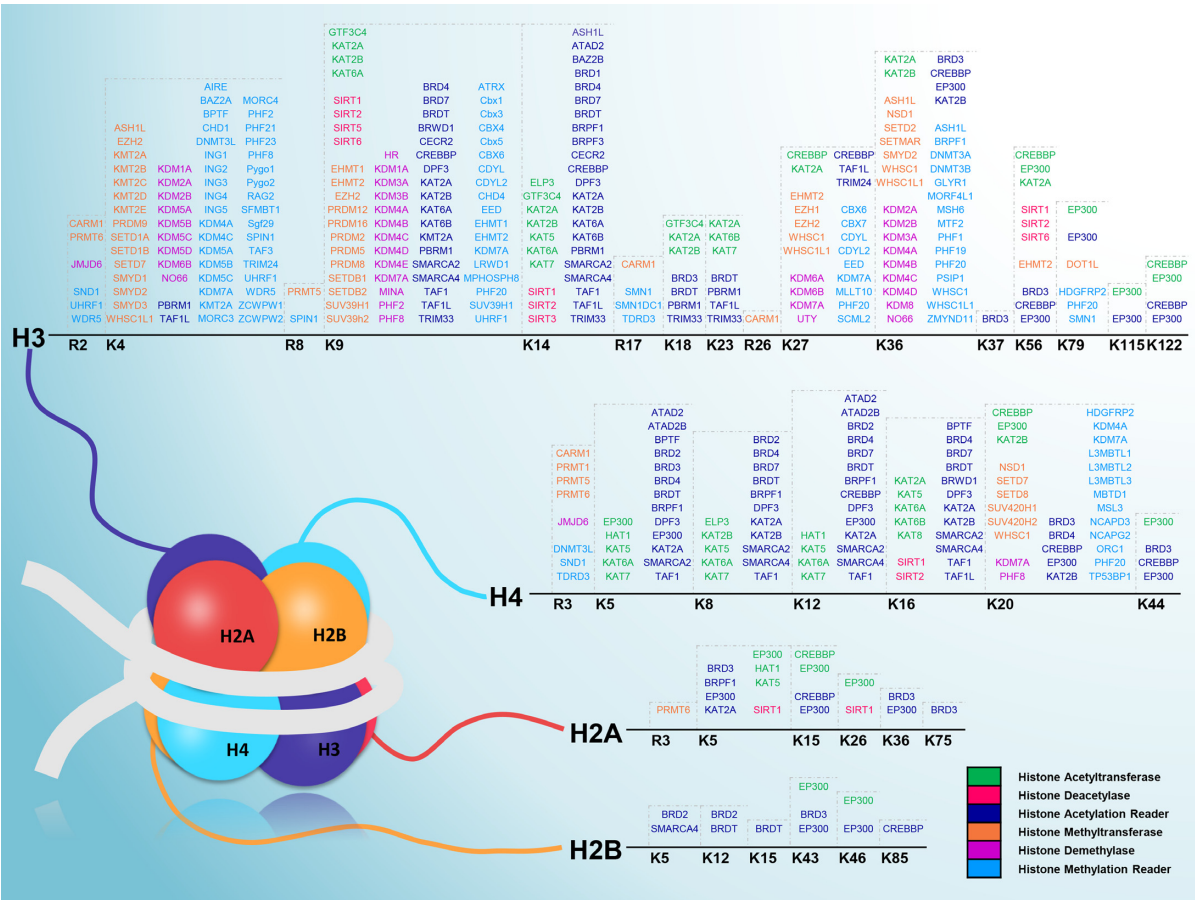


Figure 3. A snapshot of human histone modifications, containing 413 site-specific regulator-histone relations and 168 histone regulators in four core histones. The starting methionine was not counted when we referred to PTM sites in histones.

gene expression. Due to the data accumulation, a number of data resources have been developed. For example, we recently developed a database of compendium of protein lysine modification (CPLM) for up to 12 types of protein lysine modifications, containing 58 563 acetylation in 20 088 proteins and 1144 methylation sites in 745 substrates (40). More recently, we further collected and integrated 5421 methyllysines and methylarginines of 2592 proteins from the literature, and designed a computational tool of GPS-MSP for the prediction of methylation types of modified lysine and arginine residues (41). In addition, dbPTM (42), HPRD (43), PHOSIDA (44), PhosphositePlus (45) and UniProt (32) also maintained and annotated known protein acetylation and/or methylation sites.

Although more and more non-histone proteins have been identified to be acetylated or methylated, histone modifications remain to be further analyzed because the histone code is still far from fully understood. Recently, nearly 20 types of PTMs including both well-studied and new PTMs, such as phosphorylation, ubiquitination, crotonylation and succinylation, were discovered in histone variants (1,4,5). However, the identification of HATs, HDACs, HMTs, HDMs, acetyl- or methyl-binding proteins and site-specific regulator-histones relations is still an attractive challenge, since acetylation and methylation are two most predominant PTMs of histones (3,8–14,17,18,20). In

2007, 96 known histone modifying and demodifying enzymes together with 161 site-specific enzyme-histone relations were summarized and shown (17). Among four core histones, there were 56 acetylation and methylation regulators with 89 enzyme-specific sites in *H. sapiens* (17). In our study, besides the integration and identification of 20 033 histone regulators from 148 eukaryotes, we also collected 904 known site-specific regulator-histone relations of eight model organisms for acetylation and methylation. For *H. sapiens*, there have been 413 site-specific relations together with 168 histone regulators in the four core histone proteins (Figure 3). Most of these regulations were newly identified after 2007. For example, in 2009, it was reported that two human HATs, CBP and p300, acetylate histone H3 on lysine 56 (H3K56), which can be deacetylated by two human HDACs including SIRT1 and SIRT2. Such a regulatory mechanism is conserved, while *Drosophila* CBP/Nejire and Sir2 carry out acetylation and deacetylation of H3K56 in *D. melanogaster*. Acetylation of H3K56 plays an important role in chromatin assembly and DNA repair, and is significantly associated with cancers (46). Later, Tropberger *et al.* discovered that p300/CBP-mediated acetylation of H3K122 is induced by estrogen signaling and stimulates gene transcription (47). More recently, it was demonstrated that an acetylation reader, the bromodomain and PHD finger-containing protein 3 (BRPF3) forms a com-

plex with a human HAT of HBO1, and specifically acetylates H3K14 to ensure efficient activation of DNA replication (48). Thus, our results greatly expanded the integration of current knowledge on the histone code.

Taken together, our study not only generated a helpful data resource for further analysis, but also provided a better understanding of histone regulators, histone modifications and site-specific regulator-histone relations. The WERAM database will be continuously updated and improved, when new histone regulators or modifications are reported. Also, the classification and annotation information is still not integrative for several organisms, due to the relatively poor annotation of their genomes. Moreover, more species will be added when the complete proteome sets are available for the proteome-wide identification. In addition, we will further collect regulator-specific acetylation and methylation sites in non-histone proteins and provide a more useful database for protein acetylation and methylation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Funding for open access charge: National Basic Research Program (973 project) [2013CB933900]; Natural Science Foundation of China [81272578, 31671360, J1103514]; International Science & Technology Cooperation Program of China [2014DFB30020].

Conflict of interest statement. None declared.

REFERENCES

- Li, X. and Li, X.D. (2015) Chemical proteomics approaches to examine novel histone posttranslational modifications. *Curr. Opin. Chem. Biol.*, **24**, 80–90.
- 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.
- Allis, C.D. and Jenuwein, T. (2016) The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.*, **17**, 487–500.
- Zhao, Y. and Garcia, B.A. (2015) Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.*, **7**, a025064.
- Huang, H., Sabari, B.R., Garcia, B.A., Allis, C.D. and Zhao, Y. (2014) SnapShot: histone modifications. *Cell*, **159**, 458.e1.
- Yun, M., Wu, J., Workman, J.L. and Li, B. (2011) Readers of histone modifications. *Cell Res.*, **21**, 564–578.
- Patel, D.J. and Wang, Z. (2013) Readout of epigenetic modifications. *Annu. Rev. Biochem.*, **82**, 81–118.
- Smith, B.C. and Denu, J.M. (2009) Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta*, **1789**, 45–57.
- Musselman, C.A., Lalonde, M.E., Cote, J. and Kutateladze, T.G. (2012) Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.*, **19**, 1218–1227.
- Filippakopoulos, P. and Knapp, S. (2014) Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.*, **13**, 337–356.
- Chi, P., Allis, C.D. and Wang, G.G. (2010) Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer*, **10**, 457–469.
- Teperino, R., Schoonjans, K. and Auwerx, J. (2010) Histone methyl transferases and demethylases; can they link metabolism and transcription? *Cell Metab.*, **12**, 321–327.
- Greer, E.L. and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.*, **13**, 343–357.
- Liu, F., Wang, L., Perna, F. and Nimer, S.D. (2016) Beyond transcription factors: how oncogenic signalling reshapes the epigenetic landscape. *Nat. Rev. Cancer*, **16**, 359–372.
- Feinberg, A.P., Koldobskiy, M.A. and Gondor, A. (2016) Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.*, **17**, 284–299.
- Yi, X., Jiang, X.J., Li, X.Y. and Jiang, D.S. (2015) Histone methyltransferases: novel targets for tumor and developmental defects. *Am. J. Transl. Res.*, **7**, 2159–2175.
- Kouzarides, T. (2007) SnapShot: histone-modifying enzymes. *Cell*, **131**, 822.
- Kutateladze, T.G. (2011) SnapShot: histone readers. *Cell*, **146**, 842.e1.
- 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.
- Mohan, M., Herz, H.M. and Shilatifard, A. (2012) SnapShot: histone lysine methylase complexes. *Cell*, **149**, 498.e1.
- Peleg, S., Sananbenesi, F., Zovoilis, A., Burkhardt, S., Bahari-Javan, S., Agis-Balboa, R.C., Cota, P., Wittmann, J.L., Gogol-Doering, A., Opitz, L. *et al.* (2010) Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science*, **328**, 753–756.
- Britton, L.M., Gonzales-Cope, M., Zee, B.M. and Garcia, B.A. (2011) Breaking the histone code with quantitative mass spectrometry. *Expert Rev. Proteomics*, **8**, 631–643.
- Hattori, T., Taft, J.M., Swist, K.M., Luo, H., Witt, H., Slattery, M., Koide, A., Ruthenburg, A.J., Krajewski, K., Strahl, B.D. *et al.* (2013) Recombinant antibodies to histone post-translational modifications. *Nat. Methods*, **10**, 992–995.
- Yuan, Z.F., Arnaudo, A.M. and Garcia, B.A. (2014) Mass spectrometric analysis of histone proteoforms. *Annu. Rev. Anal. Chem. (Palo Alto Calif)*, **7**, 113–128.
- Zhang, Y., Lv, J., Liu, H., Zhu, J., Su, J., Wu, Q., Qi, Y., Wang, F. and Li, X. (2010) HHMD: the human histone modification database. *Nucleic Acids Res.*, **38**, D149–D154.
- Khare, S.P., Habib, F., Sharma, R., Gadwal, N., Gupta, S. and Galande, S. (2012) Histome—a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.*, **40**, D337–D342.
- Choi, J., Kim, K.T., Huh, A., Kwon, S., Hong, C., Asiegbu, F.O., Jeon, J. and Lee, Y.H. (2015) dbHiMo: a web-based epigenomics platform for histone-modifying enzymes. *Database (Oxford)*, **2015**, bav052.
- Draizen, E.J., Shaytan, A.K., Marino-Ramirez, L., Talbert, P.B., Landsman, D. and Panchenko, A.R. (2016) HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. *Database (Oxford)*, **2016**, bav014.
- 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.
- Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A. and Eddy, S.R. (2015) HMMER web server: 2015 update. *Nucleic Acids Res.*, **43**, W30–W38.
- Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
- The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
- Wang, Y., Liu, Z., Cheng, H., Gao, T., Pan, Z., Yang, Q., Guo, A. and Xue, Y. (2014) EKPD: a hierarchical database of eukaryotic protein kinases and protein phosphatases. *Nucleic Acids Res.*, **42**, D496–D502.
- Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezuk, Y. *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.

37. Deng, W., Wang, Y., Liu, Z., Cheng, H. and Xue, Y. (2014) HemI: a toolkit for illustrating heatmaps. *PLoS One*, **9**, e111988.
38. Xie, Y., Kang, R., Sun, X., Zhong, M., Huang, J., Klionsky, D.J. and Tang, D. (2015) Posttranslational modification of autophagy-related proteins in macroautophagy. *Autophagy*, **11**, 28–45.
39. Mo, F., Zhuang, X., Liu, X., Yao, P.Y., Qin, B., Su, Z., Zang, J., Wang, Z., Zhang, J., Dou, Z. *et al.* (2016) Acetylation of Aurora B by TIP60 ensures accurate chromosomal segregation. *Nat. Chem. Biol.*, **12**, 226–232.
40. Liu, Z., Wang, Y., Gao, T., Pan, Z., Cheng, H., Yang, Q., Cheng, Z., Guo, A., Ren, J. and Xue, Y. (2014) CPLM: a database of protein lysine modifications. *Nucleic Acids Res.*, **42**, D531–D536.
41. Deng, W., Wang, Y., Ma, L., Zhang, Y., Ullah, S. and Xue, Y. (2016) Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Brief. Bioinform.*, bbw041.
42. Huang, K.Y., Su, M.G., Kao, H.J., Hsieh, Y.C., Jhong, J.H., Cheng, K.H., Huang, H.D. and Lee, T.Y. (2016) dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. *Nucleic Acids Res.*, **44**, D435–D446.
43. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A. *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
44. Gnad, F., Gunawardena, J. and Mann, M. (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.*, **39**, D253–D260.
45. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V. and Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
46. Das, C., Lucia, M.S., Hansen, K.C. and Tyler, J.K. (2009) CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature*, **459**, 113–117.
47. Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Buhler, M. *et al.* (2013) Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell*, **152**, 859–872.
48. Feng, Y., Vlassis, A., Roques, C., Lalonde, M.E., Gonzalez-Aguilera, C., Lambert, J.P., Lee, S.B., Zhao, X., Alabert, C., Johansen, J.V. *et al.* (2016) BRPF3-HBO1 regulates replication origin activation and histone H3K14 acetylation. *EMBO J.*, **35**, 176–192.