Published online 2 December 2011 *Nucleic Acids Research, 2012, Vol. 40, Database issue* ***D337–D342***
doi:10.1093/nar/gkr1125

HIstome—a relational knowledgebase of human histone proteins and histone modifying enzymes

Satyajeet P. Khare[1,2], Farhat Habib[2], Rahul Sharma[2], Nikhil Gadewal[1], Sanjay Gupta[1,*] and Sanjeev Galande[2,*]

[1]Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Kharghar, Navi Mumbai 410210 and [2]Centre of Excellence in Epigenetics, Indian Institute of Science Education and Research (IISER), Pune 411021, India

Received August 15, 2011; Revised October 19, 2011; Accepted November 8, 2011

ABSTRACT

Histones are abundant nuclear proteins that are essential for the packaging of eukaryotic DNA into chromosomes. Different histone variants, in combination with their modification 'code', control regulation of gene expression in diverse cellular processes. Several enzymes that catalyze the addition and removal of multiple histone modifications have been discovered in the past decade, enabling investigations of their role(s) in normal cellular processes and diverse pathological conditions. This sudden influx of data, however, has resulted in need of an updated knowledgebase that compiles, organizes and presents curated scientific information to the user in an easily accessible format. Here, we present HIstome, a browsable, manually curated, relational database that provides information about human histone proteins, their sites of modifications, variants and modifying enzymes. HIstome is a knowledgebase of 55 human histone proteins, 106 distinct sites of their post-translational modifications (PTMs) and 152 histone-modifying enzymes. Entries have been grouped into 5 types of histones, 8 types of post-translational modifications and 14 types of enzymes that catalyze addition and removal of these modifications. The resource will be useful for epigeneticists, pharmacologists and clinicians. *HIstome: The Histone Infobase* is available online at http://www.iiserpune.ac.in/∼coee/histome/ and http://www.actrec.gov.in/histome/.

INTRODUCTION

Histones are small, highly basic nuclear proteins that associate with DNA in a specific stoichiometry to form the nucleosome, which further contributes to the formation of the chromatin fiber to package the complete genome within the nucleus. The human genome codes for more than 50 different types of histones that are expressed in a cell cycle-dependent or -independent manner. Mammalian histones have been categorized into five types; core histones H2A, H2B, H3 and H4 and a linker histone H1. Each histone category comprises of a defined repertoire of 'variants' that seem to have homo- or heteromorphous sequence variation and are expressed depending upon the cellular context. Linker histone H1 is also expressed in forms of different variants that exhibit tissue-specific expression and provide varying degree of compaction to the genome. Histones are subject to large number of reversible, enzymatic post-translational modifications (PTMs). Histones and their variants, in combination with their PTM 'code', are involved in major cellular processes like DNA damage response, X chromosome inactivation, transcriptional regulation as well as formation of an epigenetic memory (1–8). Dysregulation of such functions leads to the development of a number of diseases and syndromes (9). Hence, the information related to histone proteins that directly/indirectly affect these processes is extremely valuable for biologists.

Currently, a part of this vast information is represented by the Human Histone Modification Database (HHMD) (10), Histone (Sequence) Database (11,12), Histone Systematic Mutation Database (HistoneHits) (13) and ChromatinDB (14). The HHMD focuses on the storage and integration of histone modification information from experimental data (10). The database provides cytogenetic

*To whom correspondence should be addressed. Tel: +91 22 2740 5086; +91 22 2740 5000 5393; Fax: +91 22 2740 5085; Email: sgupta@actrec.gov.in
Correspondence may also be addressed to Sanjeev Galande, Tel: +91 20 2590 8060; Fax: +91 20 2589 9790; Email: sanjeev@iiserpune.ac.in

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

position-based and tissue-based information about histone modifications. However, since the database is primarily based on data extracted from chromatin immunoprecipitation experiments, the numbers of histone modifications covered by HHMD are limited by availability of modification specific antibodies. As a result, HHMD covers less than half of the total number of known human histone modifications with no information about modifications of linker histone H1. A large number of histone modifications show variant-specific enrichment; as a result, function of a particular type of histone modification becomes more relevant in the light of which histone variant it is expressed. Partly due to unavailability of antibodies, HHMD does not cover variant-specific information in detail. Another recent addition is the histone sequence database (11,12), which is a collection of all histones and histone-fold containing proteins from a large number of organisms including humans. The database also provides information about three-dimensional structures of histones and human histone gene complement. The database, however, does not provide detailed information about post-translational modifications of histones.

HistoneHits (13) and ChromatinDB (14) provide histone-centred information in yeast. While HistoneHits database deals with the mutation analysis of histone proteins, ChromatinDB provides genome-wide ChIP data for different histones and modifications. Although being a good model system to study epigenetic regulation by histone modification, yeast lacks the complexity shown by histone variants and their coding genes in humans.

Other than the histone-related databases, SysPTM database (15) also provides information about histone PTMs. The database covers PTM maps of histones and their variants in a number of species including humans. This database, however, provides PTM maps of only a fraction of the total human histones. This database also does not provide functional information regarding histone modifications. Multiple enzymes often modify histones in a context-dependent manner. One common disadvantage of the above mentioned databases is that they lack in information about the histone modifying enzymes. Therefore, despite the availability of these existing databases concerning histones, there remains a need for a comprehensive database that can provide a compilation of gene and protein centric functional information about histone variants, their PTMs and the modifying enzyme(s). Most importantly, interrelationship between all the above components is critical to ascertain biological relevance of the histone modifications.

Over the past decade, tremendous efforts have been directed toward understanding the epigenetic mechanisms of gene regulation. This has resulted in plethora of articles on the various molecules that are known to contribute to the epigenetic machinery (1–8). Most of these are enzymes that catalyze the addition or removal of PTMs on histones, which further dramatically affect gene regulation. On similar lines, variant-specific modifications have also become an indispensable piece of information. The amount of experimental data that exists on histones and their PTMs is ever increasing and it is now essential to draw conclusive information from all of this data. This information can also provide important insights into the study of complex diseases such as cancer.

To this end, we present *Histome*: *The Histone Infobase*, a unique relational knowledgebase encompassing detailed information about 55 histone proteins, 106 types of their distinct PTMs and 152 types of histone modifying enzymes along with their biological significance. Such comprehensive compilation of information related to histones and the epigenetic modifications is not available in any other databases available until this date.

Construction and contents

Data sources. The HIstome data and related infor mation are gathered from PubMed listed literature and publicly available UniprotKB/Swiss-Prot database (16). UniprotKB/Swiss-Prot database was selected, as it is the most comprehensive protein database with marked sites for protein modifications. Histones and modifying enzymes were searched in ‘reviewed’ entries of UniprotKB/Swiss-Prot using keyword ‘histone’ in ‘human’ species. The above results were manually curated to remove non-specific entries and add missing entries using literature. Gene related information such as symbol, name, location and GeneID was acquired from HGNC database (17). Other details were acquired from Entrez databases such as UniGene (18), OMIM (19) and RefSeq (20). Sites of histone PTMs and general information about histones and modifying enzymes was acquired through PubMed listed literature. After an exhaustive literature search we identified 55 histone variants, 106 distinct sites of their modifications and 152 modifying enzymes. Pubmed has been used to obtain information on every single protein entry (histone/enzyme) and post-translational modifications. The search was majorly carried out using specific names as well as alternative names of the proteins and coding genes. For PTMs, references were searched by using both full names and short codes. To gather information about the disease perspective for a given protein/gene/PTM, a general search was carried out in Pubmed. The resulting numbers of hits were then filtered out manually by going through relevant hits. More than 700 unique references have been listed out of which ~200 unique references are in the disease section and >500 unique references are in the database notes.

Data integration and links to external databases. Programs used to parse UniprotKB/Swiss-Prot XML files and output resulting data into MySQL tables were written in Python. Disease tables were generated manually by curating information obtained from literature. All data and information w*ere* stored in a MySQL relational database on a Linux server. Figure 1 shows a schematic layout of the database illustrating the links between different tables. Queries to the database were implemented in PHP scripts running in an Apache/PHP environment. The PHP scripting language enabled us to embed server-side code in XHTML documents. To annotate the functions of histone variants and their modifying enzymes, hyperlinks were created to UniprotKB/Swiss-Prot, HGNC, OMIM,

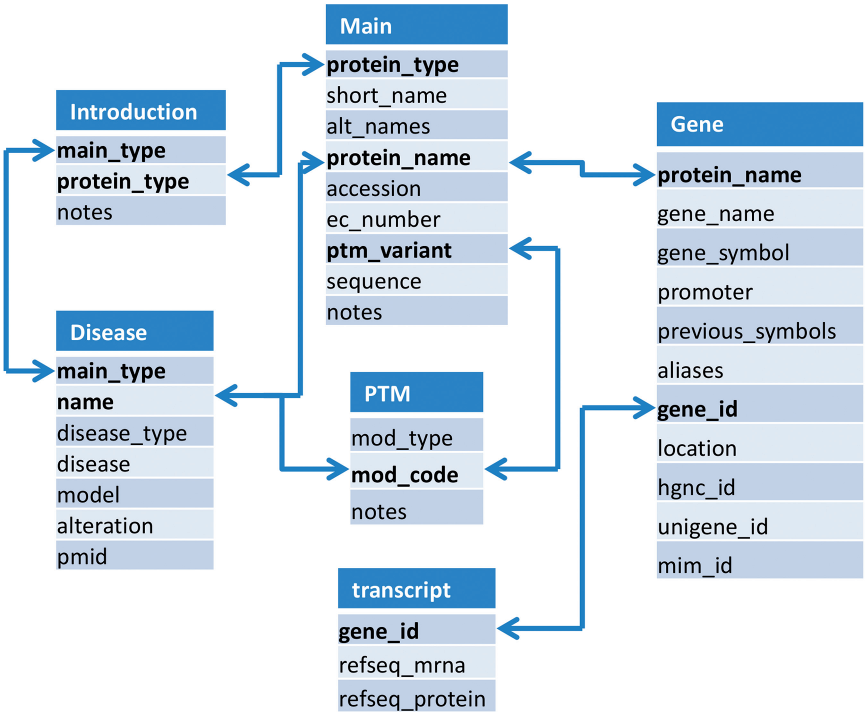


Figure 1. Organization of the histone infobase and relationships between tables. The Intro table stores information about types of histones, enzymes and PTMs. Since histone proteins are often coded by multiple non-allelic genes, and because many genes produce multiple mRNA species through alternative splicing; protein, gene and mRNA specific information has been stored in three different tables (viz. main, gene and transcript). PTM related information has been stored in a separate PTM table. Disease related information about histones, enzymes and PTMs has been stored. Tables are interlinked to other tables.

UniGene, RefSeq and other public databases. Internal hyperlinks were also created within the database pages wherever appropriate. These links greatly expand the annotation of HIstome providing related knowledge from diverse sources.

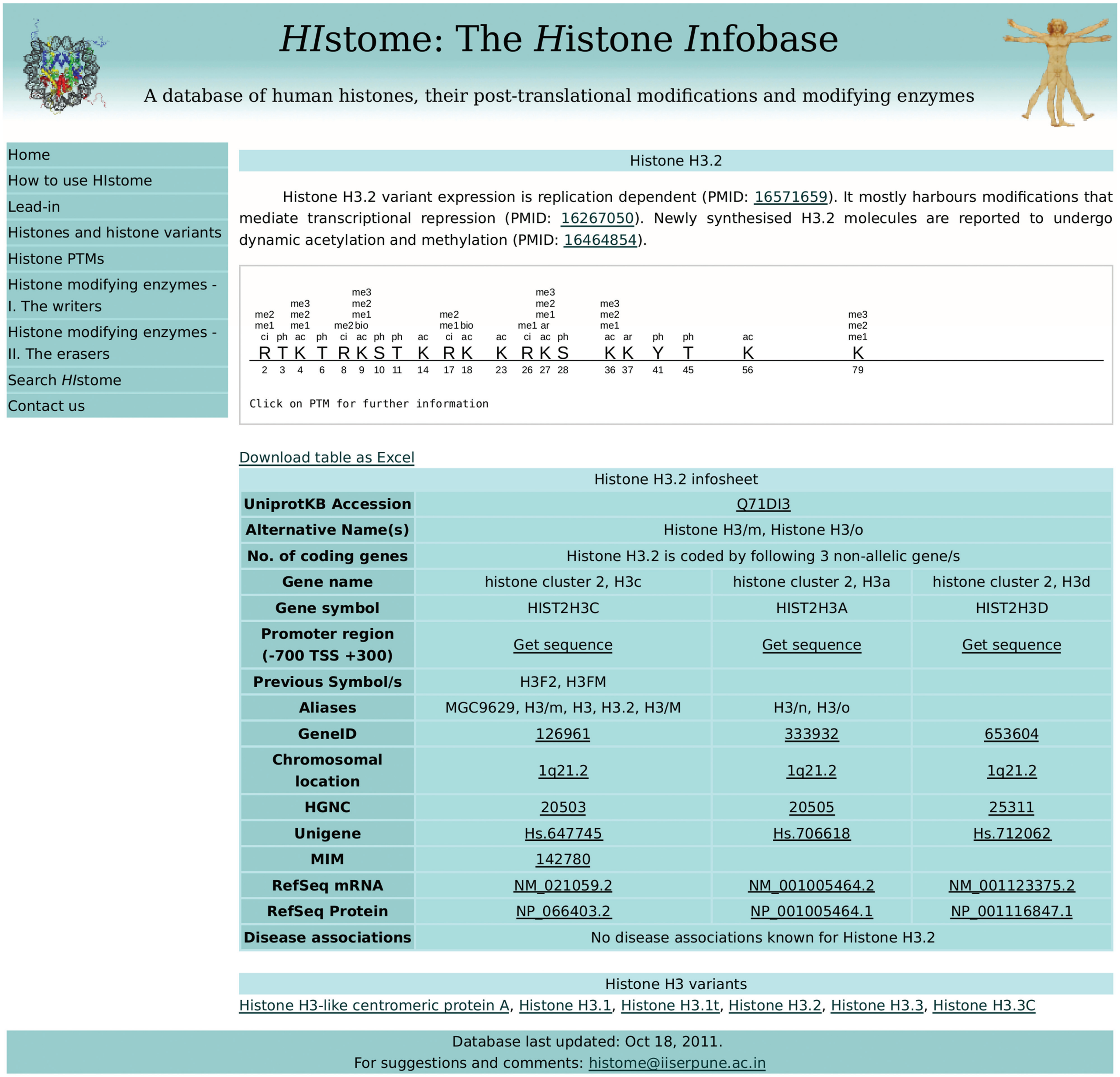
Database and website implementation. HIstome is available online and user friendly access is provided via a web interface. A detailed section has been added through side-menu on the contents of the database and how this resource can be used. A general introduction to chromatin and a detailed introduction to histones, their PTMs and modifying enzymes is available under respective menu elements. Information about different types of histones, their PTMs, various enzymes and likewise any specific entry can be retrieved in the database. The content of the database can be searched directly using any keyword(s) across the database using either a Google powered search or HIstome advanced search.

Utility and discussion

Interface and visualization. The infobase is presented using XHTML and Javascript dynamically generated using PHP with a MySQL backend. A drop-down menu bar has been provided for easy navigation of the database contents. The front page provides statistics about the total number of entries of histone proteins, their distinct PTMs and modifying enzymes in humans. A general tutorial has also been provided to explain the resource structure and it can be used in 'How to use HIstome' section. A general introduction to histone biology is provided in the 'Lead-in' section of the database. Individual records of histones, their sites of modifications and modifying enzymes can be browsed through dynamically generated menus, sub-menus and tables or via the advanced search options.

Entry information. Information about histones, PTMs and enzymes ('writers' and 'erasers') can be obtained by clicking on respective elements in the menu bar. Each of these menu elements expands into submenus, which display subcategories. Histones have been categorized into five types viz. H1, H2A, H2B, H3 and H4. Each histone page provides general information along with a table of its variants. Each variant in the table has been hyperlinked to individual variant page that provides further information (Figure 2). The variant page has also been provided with a visualization that graphically represents sites of PTMs on the histone peptide. The visualization is dynamically created with the Raphaël Javascript library. The PTMs are hyperlinked to the individual PTM pages, which can also be accessed through the menu.

Histone PTMs have been categorized into eight types depending on the type of modification and the modified amino acid, e.g. lysine acetylation, arginine methylation, serine/threonine/tyrosine phosphorylation and others. General information about each type of PTM such as



*HI*stome: The *H*istone *I*nfobase

A database of human histones, their post-translational modifications and modifying enzymes

Home
How to use HIstome
Lead-in
Histones and histone variants
Histone PTMs
Histone modifying enzymes - I. The writers
Histone modifying enzymes - II. The erasers
Search *HI*stome
Contact us

Histone H3.2

Histone H3.2 variant expression is replication dependent (PMID: 16571659). It mostly harbours modifications that mediate transcriptional repression (PMID: 16267050). Newly synthesised H3.2 molecules are reported to undergo dynamic acetylation and methylation (PMID: 16464854).

					me3								me3							
		me3			me2								me2		me3					
me2		me2			me1				me2				me1		me2					me3
me1		me1		me2	bio				me1	bio		me1	ar		me1					me2
ci	ph	ac	ph	ci	ac	ph	ph	ac	ci	ac	ac	ci	ac	ph	ac	ar	ph	ph	ac	me1
R	T	K	T	R	K	S	T	K	R	K	K	R	K	S	K	K	Y	T	K	K
2	3	4	6	8	9	10	11	14	17	18	23	26	27	28	36	37	41	45	56	79

Click on PTM for further information

Download table as Excel

Histone H3.2 infosheet

UniprotKB Accession	Q71DI3		
Alternative Name(s)	Histone H3/m, Histone H3/o		
No. of coding genes	Histone H3.2 is coded by following 3 non-allelic gene/s		
Gene name	histone cluster 2, H3c	histone cluster 2, H3a	histone cluster 2, H3d
Gene symbol	HIST2H3C	HIST2H3A	HIST2H3D
Promoter region (-700 TSS +300)	Get sequence	Get sequence	Get sequence
Previous Symbol/s	H3F2, H3FM		
Aliases	MGC9629, H3/m, H3, H3.2, H3/M	H3/n, H3/o	
GeneID	126961	333932	653604
Chromosomal location	1q21.2	1q21.2	1q21.2
HGNC	20503	20505	25311
Unigene	Hs.647745	Hs.706618	Hs.712062
MIM	142780		
RefSeq mRNA	NM_021059.2	NM_001005464.2	NM_001123375.2
RefSeq Protein	NP_066403.2	NP_001005464.1	NP_001116847.1
Disease associations	No disease associations known for Histone H3.2		

Histone H3 variants

Histone H3-like centromeric protein A, Histone H3.1, Histone H3.1t, Histone H3.2, Histone H3.3, Histone H3.3C

Database last updated: Oct 18, 2011.
For suggestions and comments: histome@iiserpune.ac.in

Figure 2. Screenshot depicting the information retrieved from a search for Histone H3.2. A visualization of all post-translational modifications on histone variant 3.2 appears at the top. The table below provides links to more information from HIstome as well as other public databases.

the donor of the functional group, molecular weight of the functional group and a list of site-specific modifications in that category can be accessed through sub-menu elements. Histone modifying enzymes are broadly categorized into ‘writers’ and ‘erasers’, those that catalyze the addition or removal of PTMs, respectively (21). Specific PTM page that provides information about particular PTM can be accessed by clicking on the PTM code. Individual writer and eraser pages can be accessed by clicking on their names or by browsing the respective menu elements as described below.

The writer enzymes have been categorized into eight types, e.g. arginine deiminases, lysine ubiquitinases, etc., depending on their catalytic activity. General information of each category of enzymes such as type of catalysis, cofactors used, etc. can be accessed by clicking on respective submenu elements that displays specific enzyme page. The enzyme page also lists various enzymes in the category and site/s of histone modification catalyzed by them. The enzymes have been hyperlinked to individual specific-enzyme pages that display manually curated information.

Description of each entry has been extracted from relevant PubMed listed literature that has been duly cited using their unique PubMed IDs (PMID). Individual histone and enzyme records also contain dynamically generated tables that provide accessions to gene, transcript and protein entries from other databases such as UniProt/Swiss-Prot, HGNC and Entrez. One Kb upstream (−700 transcription start site +300) DNA sequence has been extracted for each gene entry from UCSC genome browser and is easily accessible from the table. The PTM pages have been used to link the variant and enzyme-specific pages that assist in faster retrieval of information.

All tables that appear on the site can be downloaded in MS Excel format. The format includes the external links that appear on the page thus enabling easier downstream search. Contents of the database can be searched using a Google powered search or HIstome advanced search. The advanced search enables a targeted search for keywords in a particular table. Specifically, a user can search for a term (with wildcards) on histones, PTMs and enzymes and also filter disease associations. The results from the search lead the user directly to related detailed pages.

Over the past few years, epigenetics has emerged as one of the fastest growing areas of biomedical research. Hence, understanding various epigenetic modifications and their relationships with biological processes is of great importance. The information available on this database can be used by biology researchers to understand roles of histone modifications/variants in DNA-mediated processes such as DNA damage, transcription, cellular transformation and differentiation. Additionally, the database can also be used to understand the roles of histone modifications and the chromatin-modifying machinery toward gene activity and the maintenance and inheritance of active and inactive chromatin states. Given the significant role of the histone-modifying proteins in human disease, efforts to discover highly specific small-molecule inhibitors of these enzymes are quickly gaining momentum. Accumulating evidence suggests that histone modifications and/or components of their modification machinery are associated with the development of various human diseases including cancer, inflammation, cardiovascular and psychiatric disorders (21). Information pertaining to association of specific histone variants or histone modifications with human diseases would be also of considerable interest to researchers studying disease biology. The advanced search option available in the database can be used for mining specific information. For example query with search terms ‘diseases’ and ‘melanoma’ yields two results, one for the histone variant macro H2A.2 and another for the histone modification H3K9me3. Clicking on H3K9me3 then further provides a detailed infosheet on the enzymes and disease associations along with their Pubmed IDs.

Future development

The database content is carefully maintained separately from its presentation. This enables us to easily update the database content to reflect new information, which in turn is presented to the user. Literature searches have been planned to allow for identification and integration of new entries into the database on quarterly basis. The next major addition planned to the database is the incorporation of ‘Readers’. ‘Readers’, generally characterized by presence of certain domains that enable their binding to various PTMs, are involved in an array of cellular processes that provide meaning to the language of histone modifications (22). ‘Readers’ will also be browsable from menu as well as from entry pages of PTMs that they recognize. A module on association of histones, their PTMs and modifying enzymes with pathological conditions has been planned during expansion phase. We also plan to include entries from other species, especially model organisms, to broaden the scope of the database to a larger audience.

CONCLUSION

HIstome

The Histone Infobase is a web-based resource that provides comprehensive information about human histone proteins and their variants. It also lists and describes histone post-translational modifications and enzymes responsible for addition and removal of these PTMs from histone peptides. Each enzyme and histone entry has been provided with external links to other public databases. The database entries are cross-referenced with each other and can be browsed through menu as well as through individual entries, thus providing multiple ways to access the same information. This database will be a valuable resource for researchers as well as students working in the rapidly growing field of histone biology and epigenetic regulation.

Availability and requirement

HIstome is freely available at http://www.iiserpune.ac.in/∼coee/histome/index.php and at http://www.actrec.gov.in/histome/index.php. The database is fully functional with all standards compliant web browsers.

ACKNOWLEDGEMENTS

The authors thank Aarti Venkat, Meenakshi Sharma and Tejaswini Pachpor who helped in data mining and members of Gupta and Galande labs for user feedback. S.P.K. conceived the idea, designed the user interface, MySQL database and wrote PHP/HTML code and contributed to writing the manuscript. F.H. performed data mining and integration, wrote visualization code for graphical representation of histone modifications and participated in designing the MySQL database, contributed toward writing PHP/HTML code and manuscript. RS manually curated the data entries, carried out the literature search and wrote notes for database entries. N.G. was primary beta tester. S.Gu conceived the idea and coordinated the project. S.Ga conceived the idea, coordinated the project and wrote the manuscript. All authors have read and approved the final version of manuscript.

FUNDING

ACTREC (to Gupta Lab); ‘Centre of Excellence in Epigenetics’ grant by the Department of Biotechnology, Government of India (to Galande lab). Funding for open access charge: IISER Pune Institutional Grant.

Conflict of interest statement. None declared.

REFERENCES

1. Attikum,H.V. and Gasser,S.M. (2009) Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol.*, **19**, 207–217.
2. Bannister,A.J. and Kouzarides,T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.
3. Bonasio,R., Tu,S. and Reinberg,D. (2010) Molecular signals of epigenetic states. *Science*, **330**, 612–616.
4. Chow,J. and Heard,E. (2009) X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol.*, **21**, 359–366.
5. Koina,E., Chaumeil,J., Greaves,I.K., Tremethick,D.J. and Graves,J.A. (2009) Specific patterns of histone marks accompany X chromosome inactivation in a marsupial. *Chromosome Res.*, **17**, 115–126.
6. Oliver,S.S. and Denu,J.M. (2011) Dynamic interplay between histone H3 modifications and protein interpreters: emerging evidence for a "histone language". *Chembiochem.*, **12**, 299–307.
7. Singh,R.K. and Gunjan,A. (2011) Histone tyrosine phosphorylation comes of age. *Epigenetics*, **6**, 153–160.
8. Zhu,Q. and Wani,A.A. (2010) Histone modifications: crucial elements for damage response and chromatin restoration. *J. Cell Physiol.*, **223**, 283–288.
9. Chi,P., Allis,C.D. and Wang,G.G. (2011) Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer*, **10**, 457–469.
10. 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.
11. Mariño-Ramírez,L., Hsu,B., Baxevanis,A.D. and Landsman,D. (2006) The histone database: a comprehensive resource for histones and histone fold-containing proteins. *Proteins*, **62**, 838–842.
12. Mario-Ramrez,L., Levine,K.M., Morales,M., Zhang,S., Moreland,R.T., Baxevanis,A.D. and Landsman,D. (2011) The histone database: an integrated resource for histones and histone fold-containing proteins. *Database* (in press).
13. Huang,H., Maertens,A.M., Hyland,E.M., Dai,J., Norris,A., Boeke,J.D. and Bader,J.S. (2009) HistoneHits: a database for histone mutations and their phenotypes. *Genome Res.*, **19**, 674–681.
14. O'Connor,T.R. and Wyrick,J.J. (2007) ChromatinDB: a database of genome-wide histone modification patterns for *Saccharomyces cerevisiae*. *Bioinformatics*, **23**, 1828–1830.
15. Li,H., Xing,X., Ding,G., Li,Q., Wang,C., Xie,L., Zeng,R. and Li,Y. (2009) SysPTM: a systematic resource for proteomic research on post-translational modifications. *Mol. Cell Proteomics*, **8**, 1839–1849.
16. Consortium, Uniprot. (2010) The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
17. Seal,R.L., Gordon,S.M., Lush,M.J., Wright,M.W. and Bruford,E.A. (2011) genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.
18. Pontius,J.U., Wagner,L. and Schuler,G.D. (2003) UniGene: a unified view of the transcriptome. *The NCBI Handbook*. National Center for Biotechnology Information, Bethesda (MD).
19. Amberger,J., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2009) McKusick's online Mendelian inheritance in man (OMIM). *Nucleic Acids Res.*, **37**, D793–D796.
20. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
21. Martín-Subero,J.I. and Esteller,M. (2011) Profiling epigenetic alterations in disease. *Adv. Exp. Med. Biol.*, **711**, 162–177.
22. Tarakhovsky,A. (2010) Tools and landscapes of epigenetics. *Nat. Immunol.*, **11**, 565–568.