**“HistoneDB 2.0 - with Variants”: an integrated resource to explore histones and histone variants**

Eli J. Draizen1#, Alexey K. Shaytan1#, Leonardo Marino-Ramirez1, Paul B. Talbert2, David Landsman1\*, Anna R. Panchenko1\*

1National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA

2Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

# - authors contributed equally to this work

Corresponding authors:

Anna Panchenko ([panch@ncbi.nlm.nih.gov](mailto:panch@ncbi.nlm.nih.gov)), David Landsman ([landsman@ncbi.nlm.nih.gov](mailto:landsman@ncbi.nlm.nih.gov))

**Abstract**

Database URL: <http://www.ncbi.nlm.nih.gov/projects/HistoneDB>2.0

**1. Introduction**

Nucleosomes constitute the elementary building blocks of chromatin and play important functional roles in epigenetic regulation of transcription, replication, cell development and reprogramming. Each nucleosome core particle consists of about 147 base pairs (bp) of DNA wrapped around an octamer of histone proteins (two copies of H3, H4, H2A, H2B –) ([1](#_ENREF_1" \o "Kornberg, 1974 #2325),[2](#_ENREF_2" \o "Luger, 1997 #98)). Adjacent nucleosomes are separated by stretches of linker DNA of varying length up to about.100bp. All four types of core histones share the same “histone fold” while the sequence identity between them might not exceed 25%. The linker histone H1 usually binds to the nucleosome core and linker DNA to form “chromatosomes” and promotes further chromatin compaction. A linker histone H1 has a different fold and makes a unique set of interactions with the linker DNA([3](#_ENREF_3" \o "Harshman, 2013 #2883)). The structure of the nucleosome core as revealed by the X-ray crystallography is very conserved from throughout all eukaryotes irrespective of histone sequence variants, mutations and post-translational modifications([4](#_ENREF_4" \o "Tan, 2011 #102)). However, it has been revealed that chromatin and nucleosomes may undergo substantial conformational changes and the balance between different conformations may be shifted even by subtle changes in histone sequences ([5](#_ENREF_5" \o "Luger, 2012 #68)). Most eukaryotes have, histone variants, for some or all of the histones (H2A, H2B, H3, H4, H1). In addition, nucleosomes may employ different sets of histone variants, and post-translational modifications, which may be essential for sustaining their diverse functions and responding to the environmental stimuli ([6-8](#_ENREF_6" \o "Talbert, 2014 #2440)).

isoforms ‘’‘’ ([9](#_ENREF_9" \o "Talbert, 2010 #2613))This division is based on the history of their discovery and has its limitations. ‘Canonical’ histones in plants and animals usually encompass a variety of structurally distinct paralogs of H2A and H2B, while in many unicellular organisms, there is no special set of ‘canonical’ replication-coupled paralogs, and ‘variants’ fulfill their roles, rendering the distinction between these classes meaningless in these organisms. In animals, typically along the chromosome, lack introns, and employ a specific type of regulation at the RNA level with a stem loop structure at the 3’ end instead of polyA tail On the other hand, genes encoding histone variants are usually not clustered, have introns, and their mRNAs are regulated with polyA tails similar to the mRNAs of most genes ([10](#_ENREF_10" \o "Marzluff, 2008 #2573)) In plants, canonical histone genes lack introns, but are not clustered and the mRNAs are polyadenylated. Remarkably, more complex multicellular organisms typically have a higher number of histone variants providing a variety of different functions. Recent data are accumulating about the roles of diverse histone variants highlighting the functional links between variants and the delicate regulation of organism development. Some of the most striking examples include the importance of the H2A.Z variant in memory consolidation ([11](#_ENREF_11" \o "Zovkic, 2014 #2645)) and the modulation of the olfactory neurons life span by histone variant H2B.E in mice ([12](#_ENREF_12" \o "Santoro, 2012 #2358)).

Each histone variant has characteristic sequence and structural features that account for its specific function. The similarity between the canonical histones and the variants can be minor with only very few amino acid differences (eg. canonical H3 and H3.3) and overall conservation of most structural features. However, in some particular cases histone variants might differ from their canonical counterparts in sequence as much as different families of canonical histones differ from each other (~25% identity). Variant sequences may have shortened or extended N- and C-terminal tails and may have characteristic regions with physico-chemical properties drastically different from the canonical histones. Many of these features and their functional implications are largely unknown and/or poorly annotated. The phylogenetic origin of histone variants has been addressed in several studies, which pointed to a monophyletic origin for some variants while others, including canonical histones, were found to originate repeatedly in evolution ([13-15](#_ENREF_13" \o "Talbert, 2012 #2208)). Although some histone variants can have unique post-translational modification patterns, the majority of them remain to be characterized.

In this paper we present a database “HistoneDB2.0 – with Variants” that collects canonical histones and histone variants, their sequence, structural and functional features. This database is the successor to a previous “Histone Database” which represented a curated collection of sequences and structures of histones and non-histone proteins containing histone folds ([16-18](#_ENREF_16" \o "Baxevanis, 1996 #34)). The HistoneDB2.0 consists of two parts. First, we compiled a manually curated set of histone variants and their multiple sequence alignments with the expert annotated characteristic features and descriptions of their functions. Second, we constructed Hidden Markov Models (HMM) based on these alignments and used them together with the sequence motif identification algorithms to search a sequence of interest or all sequences from the non-redundant (NR) database given that they pass rigorous criteria established in this study. As a result, automatic annotations of histone variants can be produced. Furthermore, HistoneDB2.0 allows for comparing variants and their features to each other within the same or different species. The phylogenetic tree of histone variants offers another important evolutionary aspect of the database so that it is feasible to browse through the lineage-specific or universally conserved variants and decipher their characteristic features. The database promotes the new nomenclature for histone variants proposed recently ([13](#_ENREF_13" \o "Talbert, 2012 #2208)).

**2. Database source and contents**

The HistoneDB2.0 database consists of two complimentary parts (a) a set of manually curated and annotated variants and (b) a set of automatically extracted, classified and annotated histone sequences from non-redundant (nr) database of protein sequences maintained by NCBI. Below we describe each part of the database, its content including feature descriptions and annotations.

**2.1. Curated set of histone sequences and alignments**

histone variant ([13](#_ENREF_13" \o "Talbert, 2012 #2208)) and appended it with a set of canonical histones for a wide set of speciesthe alignment tool([19](#_ENREF_19" \o "Edgar, 2004 #2886)) en, and were in agreement with the structural alignment of the available PDB structures of histones and nucleosomes

The curated set contains histone sequences classified in total into 30 different groups representing major histone variants and canonical histones from core families H2A, H2B, H3, H4 and linker histone H1 family. The canonical sets of sequences for core histones and a generic set for H1 histone (see below) are provided as separate groups within the corresponding histone families. Note that almost no variants are available for H4 histone. Each histone type and variant has an annotation record in the database with a brief description, relevant references, structural and functional features.

We adhere to the naming convention of histone variants put forward in ref. ([13](#_ENREF_13" \o "Talbert, 2012 #2208)), while we provide alternative names on the summary page for every variant. The canonical histones are referenced as the name of histone type prefixed by “canonical”. The current version of the database focuses on indexing the major structurally distinct monophyletic clades of histone families, which according to new nomenclature are denoted with letter suffixes or prefixes (eg. H2A.Z, cenH3, etc). However, the database also includes certain variants that are denoted by number suffixes. According to ([13](#_ENREF_13" \o "Talbert, 2012 #2208)) the variants with number suffixes should be assumed to be species-specific, but in related species, where unique orthologies are clear the variants with number suffixes should correspond to related proteins. In current version of the database we opted to include groups of number suffixed variants when the sequences within each group are known to be related within a certain taxonomic span.

Below we briefly describe different histone groups indexed in the database and their main features. For brevity we do not describe the general features and functions of histone type families, although they are also available in the database. The statistics of our curated and automatically annotated sets is given in Table 1.

***Histone H2A family***

Histone H2A has the highest number of known variants (nine sequence groups indexed in our database including canonical H2A), some of which are relatively well characterized:

* H2A.X is the most common H2A variant, with the defining sequence motif ‘SQ(E/D)Φ’ (where Φ-represents a hydrophobic residue, usually Tyr in mammals). It becomes phosphorylated during the DNA damage response, chromatin remodeling, and X-chromosome inactivation in somatic cells. H2A.X and canonical H2A have diverged several times in phylogenetic history, but each H2A.X version is characterized by similar structure and function, suggesting it may represent the ancestral state.
* H2A.Z regulates transcription, DNA repair, suppression of antisense RNA, and RNA Polymerase II recruitment. Notable features of H2A.Z include a sequence motif ‘DEELD,’ a one amino acid insertion in L1-loop, and a one amino acid deletion in the docking domain. Isoform H2A.Z.2 was suggested to be driving the progression of malignant melanoma ([20](#_ENREF_20" \o "Vardabasso, 2015 #2887)). H2A can be exchange in nucleosome for H2A.Z with special remodeling enzymes ([21](#_ENREF_21" \o "Ranjan, 2015 #2904)).
* macroH2A contains a histone fold domain and an extra, long C-terminal macro domain which can bind ADP. This histone variant is used in X-inactivation and transcriptional regulation. Structures of both domains are available, but the inter-domain linker is too flexible to be crystalized.
* H2A.B (Barr body deficient variant) is a rapidly evolving mammal specific variant, known for its involvement in spermatogenesis. H2A.B has a shortened docking domain, which wraps a shorter DNA region.
* H2A.L and H2A.M variants are closely related to H2A.B, but are sufficiently less studied.
* H2A.W is a plant specific variant, that has SPKK motifs at the N-terminus, which have putative minor-groove-binding activity.
* TS H2A.1 is a mammalian testis-specific variant. It can preferentially dimerize with TS H2B.1. So far characterized only in mouse, but a similar gene in human is available. The gene is located at the end of the largest histone gene cluster.

Currently other less extensively studied H2A variants are starting to emerge such as H2A.J ([22](#_ENREF_22" \o "Shaytan, 2015 #2852)), these may be included in our database at the next update.

***Histone H2B family***

H2B histone type is known to have a limited number of variants at least in mammals, apicomplexa and sea urchins.

* TS H2B.1 is a testis-specific variant that forms subnucleosomal particles in spermatids. It can dimerize with H2A.L and TS H2A.1.
* H2B.W is involved in spermatogenesis, telomere associated functions in sperm and is found in spermatogenic cells. It is characterized by the extension of the N-terminal tail.
* subH2B participates in regulation of spermiogenesis and is found in non-nucleosomal particle in the subacrosome of spermatozoa. This variant has a bipartite nuclear localization signal.
* H2B.Z is an apicomplexan specific variant that is known to interact with H2A.Z.
* “sperm H2B” is a putative group in our database that containing sperm, early and cleavage H2B histones from sea and sand urchins and potentially is common for Echinacea.

the , it and might be included in the next update of the database.

**Histone H3 family**

Histone H3 participates of formation of H3/H4 tetramer within nucleosome and has a number of important variants present throughout eukaryotes.

* cenH3, or centromeric H3, replaces canonical H3 in centromeric nucleosomes and is essential for kinetochore formation in most eukaryotes. cenH3 typically have about 50-60% amino acid identity to canonical H3 in the histone fold domain and no conservation of its N-terminal region. It has an extended L1-loop and usually replaces Phe84 in canonical H3 with Trp, and Thr 107 with Ala, Cys, or Ser ([23](#_ENREF_23" \o "Postberg, 2010 #2888)). The formation of cenH3 nucleosomes has been controversial and may differ between species. A hemisome structure of centromeric nucleosomes in budding yeast has been proposed, with one copy of cenH3, H4, H2A and H2B and DNA forming a hemisome instead of a full nucleosome.
* H3.3 refers to a replication-independent H3, which typically differs from canonical H3 by only a few amino acids that are necessary for replication-independent assembly. H3.3 and canonical H3 diverged independently in plants, animals and ciliates. Fungi typically have H3 that undergoes both replication-coupled and replication-independent assembly, similar to H3.3, which may represent the ancestral state.
* Other H3 variants are lineage-specific, and often include germ cell or pollen variants. In our database we include TS H3.4, H3.5 and H3.Y, specific for mammals, hominid and primates respectively.

**Histone H1 family**

Histone H1 variants lack histone fold and typically have a short basic N-terminal domain, a globular winged-helix domain and a lysine- alanine-rich C-terminal domain. Their binding to nucleosome may depend on histone variant and DNA sequence([24](#_ENREF_24" \o "Cui, 2009 #2885),[25](#_ENREF_25" \o "Zhou, 2015 #2901)). They evolve faster than the core histone families and are abundant and diverse, which represents problems for their grouping and classification. Although certain H1 variants manifest replication independent or replication dependent behavior, it is difficult to name any group of sequences as canonical H1s because of high variability between species. To add to that one of the common histone variants H1.0 found across metazoa has replication independent expression as opposed to canonical core histone which are replication dependent. Hence, we opt to have in our database only several H1 histone sequence groups limited to certain taxonomic clades and a generic H1 group (genericH1), which collects various H1 sequences from a wide set of species. Currently, the following groups are present:

* Generic H1 – is a set of sequence variants across broad span of taxa not specifically related between each other.
* H1.0 is a replication independent linker histone found in animals expressed in terminally differentiated cells. Has a common monophyletic origin that can be traced back before the differentiation between protostomes and deuterostomes, very early in metazoan evolution.
* TS H1.6: Sequences collected in this group belong to (TS) H1.6 - a testis specific variant of H1 common in mammals.
* TS H1.7: Sequences collected in this group belong to (TS) H1.7 - a testis specific variant of H1 common in mammals.
* OO H1.8: Sequences collected in this group belong to (OO) H1.8 - an oocyte specific variant of H1 common in mammals.
* TS H1.9: Sequences collected here belong to (TS) H1.8 - a testis specific variant of H1 common in mammals.
* H1.10: Sequences collected here belong to H1.10 here - a vertebrate specific H1 variant.
* ScH1: A special variant of H1 found in Saccharomyces and probably other yeast species that has two globular domains. Saccharomyces cerevisiae has only one gene that encodes H1 histone (HHO1).

**2.2. Automatically extracted and annotated sets of histone sequences**

Curated alignments of histone variants were used to train Hidden Markov Models, utilizing HMMER 3.1b2 ([26](#_ENREF_26" \o "Eddy, 2011 #2889)), to create one HMM for each variant. These models were used as a part of automatic extraction and annotation pipeline. Namely, all sequences from the nr database have been classified by the HMM models (variants and canonical) and were added into the HistoneDB 2.0 as “automatically extracted sequences”. We assigned a model with a maximum HMMER score to a given sequence. The score was required to exceed a certain threshold identified as follows. For any given variant model, we calculated respective HMMER scores for all curated sequences. We then tested a binary classifier based on the score threshold that could distinguish between the variant sequences and all other sequences with 90% specificity. Specificity of retrieval was estimated based on the number of true positives (TPs; sequences correctly predicted by their native model) and false positives (FPs; incorrectly predicted sequences) found above each HMMER score cutoff. The specificity was calculated as TN/(FP+TN). Then desired cutoff value for score was obtained from the interpolated inverse curve of score cutoffs versus specificity (see Supplementary Information for details about evaluating each variant model). As can be seen in Figure X, 90% specificity threshold also results in a very high sensitivity where X% of all positive cases are classified correctly.

The HMM search algorithm was supplemented by the pattern matching in the case of H2A.X, since this variant is characterized by the presence of ‘SQ(E/D)Φ’ motif as described above. Any sequence that was classified as canonical H2A by HMMER-based algorithm was reclassified as H2A.X if it had this motif at its C-terminus.

**2.3. Histone features and annotations**

Every histone type and variant in our database has three types of manually collected annotations: (a) a brief description of the type/variant, (b) a set of structural and functional features and their locations along the sequence, (c) a list of related references. and annotations and The positions of structural and functional features along the sequence are provided with respect to the representative histone sequence of the corresponding variant. The locations of the features on other sequences or multiple sequence alignments (MSA) are inferred automatically by performing a global sequence alignment of the representative sequence with the sequence of interest or with the consensus sequence of MSA.

**3. Website overview**

The database website provides extensive functionality to: (a) browse histone variants, their annotation, features and sequences, (b) analyze phylogenetic trees of histone variants, (c) perform multiple sequence alignments of various sequences and browse them together with annotations, (d) study the taxonomic distribution of histone variants in the automatically extracted sequence set, (e) classify a user provided sequences and find the closest matches in the HistoneDB 2.0 database.

**3.1. Browsing the variants**

The front page allows choosing one of the five histone types, by selecting a color coded 3D model for each variant. After a histone type is chosen, the user is redirected to the histone type summary page that contains a list of variants with their alternate names, taxonomic spans, and sequence counts in curated and automatically extracted sequence sets. One can also see a phylogenetic tree, which shows if variants have mono- or polyphyletic origins and analyzes the relationship between different variants of the same histone type. A click on the species name within the tree will redirect the user to the “Curated sequences” (see below) tab of the respective variant with the selected sequence selected in the table. By clicking on the histone variant name the user is redirected to the histone variant summary page, which includes its description, a preview of the histone sequence (for human if available) with the highlighted features, feature legends with descriptions, and a list of related references. Both histone type and histone variant pages have four other tabs described below.

“Curated sequences” tab displays a table with a set of curated sequences with their corresponding identifiers, sequence descriptions and taxonomic classification. By clicking on the row of this table, a user can select variant’s sequence to be aligned to the canonical sequence of the same histone type. This alignment will be displayed above the table and allows to highlight the variant specific features. If one or more sequences are selected, there are options to view a multiple sequence alignment, export it in FASTA format or add it to a basket to combine sequences from different variants.

The “Curated alignments” tab depicts a multiple sequence alignment of all curated sequences for the selected histone type or variant. Alignments are annotated with the key sequence, structural and functional features, such as secondary structures, key arginines, acidic patches and features specific for a given variant.

The “Automatically extracted sequences” tab contains a list of all sequences automatically extracted (see previous section) and assigned to a given variant type. For each sequence, GI identifiers, variant’s name, taxonomy, a current GenBank description, HMMER bitscore and E-value are provided. If one or more sequences are selected, there are options to view a multiple sequence alignment or add them to a basket. For more advanced users, there is an option to view the scores for selected sequences against all HMM models (“Advanced”-“Score against all HMMs”). This option might be useful for examining if a sequence is similar to several models or if there is a classification error. It is also possible to filter the sequences by taxonomy, sequence headers, sequence motifs, GI identifiers, and sequences found in RefSeq.

The last tab “Inferred Taxonomy Distribution” allows browsing the taxonomy of automatically extracted sequences and viewing the corresponding classification of a sequence within the selected taxa with respect to all database models. The score represents the average HMM score of sequences from a given taxa. Scores are calculated for rank 'order,' the leaves of this tree, and it's parents scores/colors are interpolated by averaging the color hues of its children.

**3.2. Database search options**

for sequence entries in the databasea an . Both are accessible through the field and button in the header of the website. searchoptionopportunitiesThe quick search field is also implemented on the top of the table while viewing any list of sequences. In this case the search is limited to the content of the respective table. Another helpful feature of the website is “Basket”. While viewing a set of sequences, the user can mark the entries of interest in the table and press the button “Add to basket”. The cumulative list of sequences can be then viewed at the basket page accessible from the website header.

**3.3. Custom sequence annotation**

A user can enter a sequence in FASTA format to annotate it through the “Analyze Your Sequence” utility. A sequence is classified using HMMER against all variant models and by running BLAST against all curated and automatically extracted sequences in the HistoneDB 2.0. The results are shown with the most likely variant or variants indicated in green, a comparison between the query sequence and the canonical sequence of the same or similar species determined by BLAST along with the feature annotations. HMMER scores against all models are listed in the window.

**4. Methods, internal structure and software**

The HistoneDB 2.0 is written in Django, a high-level Python Web Framework, with a MySQL backend. The project has two applications, ‘browse,’ and ‘djangophylcore.’ Browse contains the HistoneDB models (equivalent to database tables), views (python functions to render each page), and templates (HTML files). Djangophylocore is a previously developed django application to store the NCBI taxonomy database in a Django relational database using an algorithm similar to Modified Preorder Tree Traversal([27](#_ENREF_27)). The website layout is based on Twitter Bootstrap, with four important pages: Main browse of all histone types, Individual histone type browse, Variant browse, Analyze, and Search.

Hidden Markov Model construction and search relies on HMMER 3.1b2 ([26](#_ENREF_26" \o "Eddy, 2011 #2889)). Phylogenetic tree is displayed using jsPhyloSVG software ([28](#_ENREF_28" \o "Smits, 2010 #2891)), the tree is created by aligning all curated sequences for a given histone type using MUSCLE v3.8.31 ([19](#_ENREF_19" \o "Edgar, 2004 #2886)) and by further applying the Neighbor-Joining procedure implemented in CLUSTALW 2.1 ([29](#_ENREF_29" \o "Larkin, 2007 #2892)). The trees were converted to PhyloXML using BioPython ([30](#_ENREF_30" \o "Cock, 2009 #2893)) and were edited to add colors and variant and taxonomy labels in jsPhyloSVG. Alignments are displayed using BioJS MSA Viewer([31](#_ENREF_31" \o "Gomez, 2013 #2894)).

**5. Conclusions**

Despite the considerable progress in sequencing and understanding the functions of histone variants, many of them remain poorly annotated in public databases. Moreover, the specific molecular mechanisms of variants’ action, deposition and eviction are still unknown. One of the reasons is that the histone variants are highly specific and at the same time multi-functional and context dependent. The importance of studying the histone variants is difficult to overestimate; they are involved in regulation of many cellular processes and represent emerging key players in cancer ([20](#_ENREF_20" \o "Vardabasso, 2015 #2887),[32](#_ENREF_32" \o "Maze, 2014 #2315)). To analyze all known histone variants, compare them and interpret their functions would be challenging and time consuming. To fulfill this goal we designed the HistoneDB2.0 in order: -to organize histones by variant; - to provide reference alignments for each variant; - to offer curated annotations of variant specific features; - to understand how variants evolved; - to find orthologs of variants in other species; and finally, - to promote the new histone variant nomenclature ([13](#_ENREF_13" \o "Talbert, 2012 #2208)). This database allows finding variant sequences from different organisms and comparing histone variants and corresponding canonical histones. This in turn will help to understand the origin of functional specificity of variants and to guide the 3D modeling of variant nucleosomes. The database can be easily extended to include new variants and annotate them based on the similarity to the existing annotated variants.

**Acknowledgements**

This work was supported by the Intramural Research Program of the National Library of Medicine, NIH. AS was supported in part by the US-Russia Collaboration in the Biomedical Sciences NIH visiting fellows program. We would like to thank Franco Simonetti for discussions about Django and the IT support team at NCBI for the troubleshooting the server related issues.

Fig 1. The Nucleosome core particle. H2A is yellow, H2B is red, H3 is blue, and H4 is green. H1 is not shown. Figure created by

Figures: slide#4, slide#6, #7, phylogenetic tree, ROC curve, database schema

**References**

1. Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868-871.

2. Luger, K., Mader, A.W., Richmond, R.K.*, et al.* (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251-260.

3. Harshman, S.W., Young, N.L., Parthun, M.R.*, et al.* (2013) H1 histones: current perspectives and challenges. *Nucleic acids research*, **41**, 9593-9609.

4. Tan, S., Davey, C.A. (2011) Nucleosome structural studies. *Current opinion in structural biology*, **21**, 128-136.

5. Luger, K., Dechassa, M.L., Tremethick, D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Bio*, **13**, 436-447.

6. Talbert, P.B., Henikoff, S. (2014) Environmental responses mediated by histone variants. *Trends in cell biology*.

7. Santoro, S.W., Dulac, C. (2015) Histone variants and cellular plasticity. *Trends in genetics : TIG*, **31**, 516-527.

8. Turinetto, V., Giachino, C. (2015) Histone variants as emerging regulators of embryonic stem cell identity. *Epigenetics : official journal of the DNA Methylation Society*, **10**, 563-573.

9. Talbert, P.B., Henikoff, S. (2010) Histone variants--ancient wrap artists of the epigenome. *Nature reviews. Molecular cell biology*, **11**, 264-275.

10. Marzluff, W.F., Wagner, E.J., Duronio, R.J. (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature reviews. Genetics*, **9**, 843-854.

11. Zovkic, I.B., Paulukaitis, B.S., Day, J.J.*, et al.* (2014) Histone H2A.Z subunit exchange controls consolidation of recent and remote memory. *Nature*.

12. Santoro, S.W., Dulac, C. (2012) The activity-dependent histone variant H2BE modulates the life span of olfactory neurons. *eLife*, **1**, e00070.

13. Talbert, P.B., Ahmad, K., Almouzni, G.*, et al.* (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenetics & chromatin*, **5**, 7.

14. Eirin-Lopez, J.M., Gonzalez-Romero, R., Dryhurst, D.*, et al.* (2009) Long-Term Evolution of Histone Families: Old Notions and New Insights into Their Mechanisms of Diversification Across Eukaryotes. *Evolutionary Biology: Concept, Modeling, and Application*, 139-162.

15. Malik, H.S., Henikoff, S. (2003) Phylogenomics of the nucleosome. *Nature structural biology*, **10**, 882-891.

16. Baxevanis, A.D., Landsman, D. (1996) Histone Sequence Database: a compilation of highly-conserved nucleoprotein sequences. *Nucleic acids research*, **24**, 245-247.

17. Sullivan, S., Sink, D.W., Trout, K.L.*, et al.* (2002) The Histone Database. *Nucleic acids research*, **30**, 341-342.

18. Marino-Ramirez, L., Levine, K.M., Morales, M.*, et al.* (2011) The Histone Database: an integrated resource for histones and histone fold-containing proteins. *Database : the journal of biological databases and curation*, **2011**, bar048.

19. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792-1797.

20. Vardabasso, C., Gaspar-Maia, A., Hasson, D.*, et al.* (2015) Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. *Molecular cell*, **59**, 75-88.

21. Ranjan, A., Wang, F., Mizuguchi, G.*, et al.* (2015) H2A histone-fold and DNA elements in nucleosome activate SWR1-mediated H2A.Z replacement in budding yeast. *eLife*, **4**.

22. Shaytan, A.K., Landsman, D., Panchenko, A.R. (2015) Nucleosome adaptability conferred by sequence and structural variations in histone H2A-H2B dimers. *Current opinion in structural biology*, **32C**, 48-57.

23. Postberg, J., Forcob, S., Chang, W.J.*, et al.* (2010) The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC evolutionary biology*, **10**, 259.

24. Cui, F., Zhurkin, V.B. (2009) Distinctive sequence patterns in metazoan and yeast nucleosomes: implications for linker histone binding to AT-rich and methylated DNA. *Nucleic acids research*, **37**, 2818-2829.

25. Zhou, B.R., Jiang, J., Feng, H.*, et al.* (2015) Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Molecular cell*, **59**, 628-638.

26. Eddy, S.R. (2011) Accelerated Profile HMM Searches. *PLoS computational biology*, **7**, e1002195.

27. Ranwez, V., Clairon, N., Delsuc, F.*, et al.* (2009) PhyloExplorer: a web server to validate, explore and query phylogenetic trees. *BMC evolutionary biology*, **9**, 108.

28. Smits, S.A., Ouverney, C.C. (2010) jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. *PloS one*, **5**, e12267.

29. Larkin, M.A., Blackshields, G., Brown, N.P.*, et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.

30. Cock, P.J., Antao, T., Chang, J.T.*, et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422-1423.

31. Gomez, J., Garcia, L.J., Salazar, G.A.*, et al.* (2013) BioJS: an open source JavaScript framework for biological data visualization. *Bioinformatics*, **29**, 1103-1104.

32. Maze, I., Noh, K.M., Soshnev, A.A.*, et al.* (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nature reviews. Genetics*, **15**, 259-271.