**“HistoneDB 2.0 - with Variants”: a curated database for the analysis of histone isoforms, their features and evolution**

Eli J. Draizen1#, Alexey K. Shaytan1#, Leonardo Marino-Ramirez2, Paul B. Talbert3, David Landsman2\*, Anna R. Panchenko1\*

1Computational Biophysics Group, Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, MSC 6075, Bethesda, MD 20894-6075, USA

2Bioinformatics of Chromatin Structure Group, Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, MSC 6075, Bethesda, MD 20894-6075, USA

3Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

# - authors contributed equally to this work

Corresponding authors:

David Landsman ([landsman@ncbi.nlm.nih.gov](mailto:landsman@ncbi.nlm.nih.gov)), Anna Panchenko ([panch@ncbi.nlm.nih.gov](mailto:panch@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 and cell development and reprogramming. A nucleosome core particle wraps around 147 bp of “core” nucleosomal DNA around an octamer of histone proteins (H3, H4, H2A, H2B – two copies of each) ([1](#_ENREF_1" \o "Kornberg, 1974 #2325),[2](#_ENREF_2" \o "Luger, 1997 #98)). Adjacent nucleosomes are separated by stretches of linker DNA. 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 promote 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 X-ray crystallography is very conserved from through eukaryotes irrespective of histone sequence variants, mutations and post-translational modifications([4](#_ENREF_4" \o "Tan, 2011 #102)). However, it is being appreciated that during its functioning chromatin and nucleosomes undergo substantial conformational changes and the balance between different conformations may be shifted even by subtle alteration of histone sequence([5](#_ENREF_5" \o "Luger, 2012 #68)). In most eukaryotes, there are paralogs, or ‘histone variants’, for some or all of the histone proteins (H2A, H2B, H3, H4, H1). Nucleosomes may employ different sets of histone variants, and post-translational modifications, which may be essential to perform their diverse functions and respond to environmental stimuli. ([6](#_ENREF_6" \o "Talbert, 2014 #2440)).

paralogs ‘’‘’ ([7](#_ENREF_7" \o "Talbert, 2010 #2613))This division is based on the history of their discovery, however, 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 role, 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([8](#_ENREF_8" \o "Marzluff, 2008 #2573)) In plants, canonical histone genes lack introns, but they 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 accumulating about the role of diverse histone variants, highlights striking functional links between variants and delicate regulation of organism development and functioning. Some of the most striking examples include the importance of H2A.Z histone exchange in memory consolidation([9](#_ENREF_9" \o "Zovkic, 2014 #2645)) and modulation of olfactory neurons life span by histone variant H2B.E in mice([10](#_ENREF_10" \o "Santoro, 2012 #2358)).

Each histone variant has characteristic sequence and structural features which account for its specific function. The similarity between canonical histones and other histone variants can be very substantial with the only very few amino acid differences (eg. canonical H3 and H3.3) and with overall conservation of most structural features. However, in some particular cases histone variants might differ from canonical counterparts in sequence as much as different families of canonical histones differ from each other (~25% identity). Variant sequence may have shortened or extended 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 or poorly annotated. The phylogenentic origin of histone paralogs 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([11-13](#_ENREF_11" \o "Talbert, 2012 #86)). Although some histone variants can have unique post-translational modification patterns, the majority of them remain to be characterized for these modifications.

In this paper we present a database “HistoneDB2.0 – with Variants” that collects canonical histones and histone variants, their sequence, and 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 ([14-16](#_ENREF_14" \o "Baxevanis, 1996 #34)). The HistoneDB2.0 consists of two parts. First, we compile a manually curated set of histone variants and their multiple alignments with the expert annotated characteristic features and descriptions of their functions. Second, we construct profile Hidden Markov Models (HMM) based on these alignments and use them together with motif-based identification to search any 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 are produced. Moreover, HistoneDB2.0 allows easy comparison of variants and their features to each other within the same or different species. The phylogenetic tree of histone variants presents 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 ([11](#_ENREF_11" \o "Talbert, 2012 #86)).

**2. Database source and contents**

The HistoneDB2.0 database sequence set consistes of two complimentary parts (a) a set of manually curated and annotated sequence variants and (b) a set of automatically extracted, classified and annotated histone sequences from the NCBI non-redundant database of all known sequences. Below we describe each sequence set as well as the contents of the available feature descriptions and annotations in the database.

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

([17](#_ENREF_17" \o "Edgar, 2004 #2886)) , and were in agreement with the structural alignment of the available PDB structures of histones and nucleosomes

The curated set of HistoneDB contains histone sequence classified in total into 25 different subsets, each representing a major histone variant or a canonical sequence subset. These subsets are grouped according to **histone type** family: core histones H2A, H2B, H3, H4 and linker histone H1. Note that almost no non-canonical variants are available for H4 histone. Each histone type and variant has an annotation record in the database with a brief description, relevant references and structural and functional features. Below we briefly describe different histone variants indexed in our database and their main features. For brevity we do not describe the general features and functions of the histone type families, but these are also available in our database. The statistics of our curated and automatically annotated sets if given in Table 1.

***Histone H2A family***

Histone H2A has the highest number of known variants (seven sequence sets 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)), 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 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 shown to be driving the progression of malignant melanoma ([18](#_ENREF_18" \o "Vardabasso, 2015 #2887)).
* macroH2A contains a histone fold domain and an extra 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 spermiogenesis. 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.

We note that currently other less extensively studied H2A variants are starting to emerge such as TS H2A.1, H2A.J ([19](#_ENREF_19" \o "Shaytan, 2015 #2852)), these may be included in our database at a next update.

***Hisotne H2B family***

At the current level of knowledge the H2B histone type is known to have a limited number of variants in mammals and apicomplexa.

* TS H2B.1 is a testis-specific variant that forms subnucleosomal particles in spermatids. It can dimerize with H2A.L.
* H2B.W is involved in spermiogenesis, 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 in an apicomplexan specific variant that is known to interact with H2A.Z

, but it is unclear if it has orthologs in other species, and has a set of unique features at the protein level across species. It might be included in the next update of the database.

**Histone H3 family**

* cenH3, or centromeric H3, replaces canonical H3 in centromeric nucleosomes and is essential for kinetochore formation in most eukaryotes. cenH3s typically have only about 50-60% amino acid identity to canonical H3 in the histone fold domain and no conservation of the N-terminal. cenH3s have an extended L1-loop and usually replace Phe84 in canonical H3 with Trp, and Thr 107 with Ala, Cys, or Ser ([20](#_ENREF_20" \o "Postberg, 2010 #2888)). The structure 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 molecule each of cenH3, H4, H2A and H2B and DNA forming a hemisome instead of a full nucleosome.
* H3.3 refers to replication-independent H3s, which typically differ from canonical H3s by only a few amino acids that are necessary for replication-independent assembly. H3.3 and canonical H3s diverged independently in plants, animals and ciliates. Fungi typically have H3s that undergo both replication-coupled and replication-independent assembly, similar to H3.3s, 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 H3.1, H3.2 and H3.Y.

**Histone H1 family**

H1 variants lack histone folds and typically have a short basic amino-terminal domain, a globular winged-helix domain and a lysine-rich carboxy-terminal domain. They evolve faster than the core histone families and are abundant and diverse. Many seem to have redundant functions, while some show germline specificity. We currently include ……

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

Curated alignments of histone variants were used to train Hidden Markov Models, using HMMER 3.1b2 ([21](#_ENREF_21" \o "Eddy, 2011 #2889)), creating one HMM for each variant. These models were further used as a part of automatic extraction and annotation pipeline.

All sequences from the non-redundant (nr) database of NCBI have been classified by our HMM models (variants and canonical) and were added into the HistoneDB as “automatically extracted sequences” set. We assigned a model with a maximum HMMer score for a given sequence, in addition this score was required to cross a certain threshold identified as follows. For any given model, curated sequences that were used to construct its curated alignment were treated as positives and curated sequences of other models as negatives. Then we performed HMMsearch of the database of all models against a set of curated sequences and calculated the specificity of retrieval of each sequence by its native model. 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 score HMMER score cutoff. The specificity was calculated as TN/(FP+TN). Then we interpolated an inverse curve of score cutoffs versus specificity, and obtained the score cutoff value corresponding to the specificity of 95%. Note that different models had different cutoff scores (see Supplementary Information for details about evaluating each variant model). As can be seen on Figure X, 95% 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 pattern matching in case of H2A.X, since this variant is characterized on a functional basis by the presence of ‘SQ(E/D)Φ motif as described earlier. Any sequence that was classified as canonical H2A by HMM based algorithm was reclassified as H2A.X if it had this defining motif at its C-terminus.

This classification algorithm can be also applied to classify any sequence of interest (see section 3.4. Custom sequence annotation).

The statistics of our automatically annotated set of histone sequences in shown in Table 1.

**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) set of structural features along the sequence with their description and location, (c) set of related references. and annotations and Internally, the positions of structural and functional features along the sequence are given with respect to a one representative histone sequence of the corresponding variant. The position of the corresponding features on other sequences or multiple sequence alignments (MSA) are derived automatically by performing a global alignment of the representative sequence with the sequence of interest or consensus sequence of the MSA.

**3. Web-site organization**

The database web-site provides extensive functionality to (a) browse histone variant sets, their annotation, features and sequences, (b) browse phylogenetic trees of histone variants, (c) perform multiple sequence alignments of various sequences and browse them together with annotations, (d) browse taxonomic distribution of histone variants in the automatically extracted sequence set, (e) classify a user provided sequences and find closest matches in our 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 use 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 in the list 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 highlighted features, a feature legend with descriptions, and a list of related references. Both histone type and histone variant pages have four other tabs described below.

“Curated sequences” displays 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 a variant sequence to be aligned to the canonical sequence of the same histone type, which will be displayed above the table. This 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 to a basket to combine sequences from different variants.

“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 patch 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, GenBank identifier, 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 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 was a classification error. It is also possible to filter the sequences by taxonomy, GenBank headers, sequence motifs, GI number, and sequences found in RefSeq.

The last tab “Inferred Taxonomy Distribution” allows browsing the taxonomy of automatically extracted sequences and view 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 extrapolated by averaging the colors of its children.

**3.2. Database search and basket**

for sequence entries in the database, both are accessible through the field and button in the header of the web-site. 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 web-site is the “Basket”. While viewing a set of sequences the user can tick mark the entries of interest in the table and press the button “Add to basket”. The cumulative list of sequences can be then view at the basket page accessible from the web-site header. The sequence can be then exported or an MSA can be requested.

**3.3. Custom sequence annotation**

A user can enter a sequence in FASTA format to classify through the “Analyze Your Sequence” page accessible through the web-site header. A sequence is classified using HMMER against all variant models and by running BLAST against all curated and automatically extracted sequences in the HistoneDB. The results are shown first with the most likely variant or variants, a comparison between the analyzed sequence and the canonical of the similar species determined by BLAST along with the new feature annotations. HMMER scores against all models are listed in modal 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([22](#_ENREF_22" \o "Ranwez, 2009 #2890)). 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 Models construction and search relies on HMMER 3.1b2 ([21](#_ENREF_21" \o "Eddy, 2011 #2889)). Phylogenetic tree is displayed using jsPhyloSVG software ([23](#_ENREF_23" \o "Smits, 2010 #2891)), the tree is created by aligning all curated sequences for a given histone type using MUSCLE v3.8.31 ([17](#_ENREF_17" \o "Edgar, 2004 #2886)) and by further applying the Neighbor-Joining procedure implemented in CLUSTALW 2.1 ([24](#_ENREF_24" \o "Larkin, 2007 #2892)). The trees were converted to PhyloXML using BioPython ([25](#_ENREF_25" \o "Cock, 2009 #2893)) and were edited to add colors and variant and taxonomy labels in jsPhyloSVG. Alignments are displayed using BioJS MSA Viewer([26](#_ENREF_26" \o "Gomez, 2013 #2894)).

**5. Conclusions**

Despite considerable progress in sequencing and understanding the functions of histone variants, many of them remain poorly annotated. 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 ([18](#_ENREF_18" \o "Vardabasso, 2015 #2887),[27](#_ENREF_27" \o "Maze, 2014 #2315)). To analyze all known histone variants, compare them and interpret their functions would be challenging and timely. To fulfill this goal we designed the HistoneDB2.0: 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([11](#_ENREF_11" \o "Talbert, 2012 #86)). This database allows finding and comparing variant sequences from different organisms or between histone variants and corresponding canonical histones. This in turn will helps 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. Talbert, P.B., Henikoff, S. (2010) Histone variants--ancient wrap artists of the epigenome. *Nature reviews. Molecular cell biology*, **11**, 264-275.

8. 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.

9. 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*.

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

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

12. 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.

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

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

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

16. 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.

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

18. 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.

19. 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.

20. 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.

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

22. 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.

23. 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.

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

25. 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.

26. 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.

27. 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.