Histone Isoforms and the Oncohistone Code

Andrew Flaus[1,4], Jessica A. Downs[2,4], and Tom Owen-Hughes[3,4]

[1]Centre for Chromosome Biology, Biochemistry, School of Natural Sciences, National University of Ireland, Galway, Ireland
[2]Epigenetics and Genome Stability Team, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK
[3]Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland
[4] Equal contribution

Corresponding authors: Downs, Jessica A. (Jessica.Downs@icr.ac.uk), Owen-Hughes, Tom (t.a.owenhughes@dundee.ac.uk)

Short title: Histone Isoforms and the Oncohistone Code
Keywords: Oncohistones; histone isoforms; epigenetics; cancer biology; chromatin

Abstract

Recent studies have highlighted the potential for missense mutations in histones to act as oncogenic drivers, leading to the term 'oncohistones'. While histone proteins are highly conserved, they are encoded by multigene families. There is heterogeneity among these genes at the level of the underlying sequence, the amino acid composition of the encoded histone isoform, and the expression levels. One question that arises, therefore, is whether all histone-encoding genes function equally as oncohistones. In this review, we consider this question and explore what this means in terms of the mechanisms by which oncohistones can exert their effects in chromatin.

Introduction

Hierarchical chromatin packaging by nucleosome building blocks based on an octamer of the four canonical core histone types H2A, H2B, H3 and H4 is a characteristic of eukaryotes. This chromatin is the substrate for genome transactions, and histones make multifaceted contributions to gene expression and genome stability as well as aberrant processes leading to cancer. The high homology of histone proteins in animals, plants and fungi has led to a widespread assumption that they are among the most conserved proteins in eukaryotes [1] so canonical nucleosomes are often approximated as homogenous tuna cans.

In fact, canonical histones are encoded by multigene families that show microheterogeneity at the protein level and even larger variations in gene sequence. For example, the 17, 18, 15 and 15 human genes assigned as canonical H2A, H2B, H3 and H4 encode 11, 15, 4 and 2 distinct polypeptides respectively [2-4]. Although these protein isoforms typically differ pairwise by only 1-2 residues, the variations often involve post-translationally modifiable or charged sidechains. This demonstrates that histone protein variation is possible.

There is a substantial demand for histones during replication. Using simple estimates of cell parameters [5] one can estimate that packaging an extra human diploid genome during a 10 hour S phase requires sustained production of ~7000 histones per second by ~10% of cellular ribosomes. Rapidly supplying sufficient transcripts to load 10^5 ribosomes on demand is likely to require several histone gene copies [6]. Histone production is actively regulated at multiple levels [6,7] because insufficient supply can be a constraint for S phase progression resulting fork stalling [8,9] whereas overaccumulation can lead to “histone stress” resulting in aneuploidy and genome duplications [10].

All histones are not created equal, or equally

The canonical histones that participate in most nucleosomes are expressed in synchrony with replication, are encoded by multicopy gene families, and their transcripts are not polyadenylated. In contrast, variant histones are expressed independently of replication, are encoded by one or two independent genes, and are incorporated in at most a few percent of nucleosomes at restricted genomic locations for particular functions. A few single copy histone genes show strong tissue-specific expression, often in testes. An exception to this simple distinction is H2AX, which is the product a single copy *H2AFX* gene and can contribute

10% or more of total H2A distributed throughout the genome due to expression throughout the cell cycle that is elevated during replication.

In humans, over 60 canonical histone genes are clustered in the major histone locus HIST1 on chromosome 6 at the edge of the major histocompatibility cluster (MHC), with another 12 in HIST2 on chromosome 1 (Figure 1A). A further 5 genes are found in two loci referred to as HIST3 and HIST4. Marzluff and colleagues introduced a gene nomenclature in 2002 that combined the cluster, histone type and a sequential order identifier letter if appropriate (*e.g. HIST1H2BD* or *HIST3H3*) [2]. This became unwieldy so in 2019 the HUGO Gene Nomenclature Committee (HGNC) adopted a new nomenclature with genes in HIST1 and HIST2 taking the histone type, a C for clustered, and an arbitrary number with its origins in the location order (e.g. *HIST1H2BD* became *H2BC5*). Genes in HIST3 and HIST4 include a hyphen and an arbitrary number (e.g. *HIST3H3* became *H3-4*). In contrast, histone variant genes are mostly named with their type, an F for family, and an arbitrary letter or number (*e.g. H2AFX* and *H3F3A* encoding H2A.X and H3.3), and have not been renamed. Human histone nomenclature remains a challenge for users to navigate because some databases have not yet updated to the new HGNC nomenclature or have seen frequent revisions.

The regulation of histone gene isoform expression within the complex interspersed arrangements of HIST1 and HIST2 is poorly understood. Progress has been hindered by the fact that canonical histone transcripts are not polyadenylated, compromising the usefulness of most data which is generated using polyA selection. Nevertheless, transcriptomic data generated without polyA selection reveals that 4-6 canonical genes of each core histone type appear to be responsible for >75% of transcripts (Figures 1B and 2). Furthermore, the expression levels of gene isoforms vary in cells derived from different tissues (Figure 1B).

Oncohistones: When good histones go bad

Notwithstanding microheterogeneity, histones are highly conserved at the protein level so mutations are anticipated to have functional consequences. Large-scale genome and exome sequencing enables unbiased surveying of genetic alterations, and has uncovered recurrent missense mutations in specific histone genes linked to cancers [11-13]. “Oncohistones” are histone gene alleles that act as drivers of cancers, although how mutations to specific gene isoforms within multigene families exert dominant phenotypes is unclear.

Mutation of the H3.3 variant isoform H3-3A giving a lysine to methionine change was identified in paediatric glioblastomas [14]. These mutations act in a dominant fashion to impede methylation of H3K27 by the EZH2 subunit of the Polycomb Repressor complex (PRC2) which in turn results in loss of transcriptional silencing [15,16]. Mutations are also observed to affect additional sites that are subject to post-translational modifications [16]. The properties of this class of mutants correlates with phenotypes of mutations in the enzymes that read, write and erase modifications at the residues [17], consistent with the histone mutations exerting their effects via the signalling pathways dependent on the modifications.

A second group of recurrent missense mutations linked to cancers alters residues in the histone tails or in residues affecting nucleosome structure and stability [11,13,18]. These include Swi/Snf Independent (SIN) mutations and residues that affect prominent charged surfaces such as the acidic patch that interact with a broad range of chromatin binding proteins [11].

This indicates that mutations in histone genes can exert functional effects via a variety of mechanistic pathways. While many of the oncohistone examples identified so far are derived from H3 encoding gene isoforms or variants, cancer-associated point mutations are observed in all 4 core histones, linker histones and their variants (Figure 2).

An instructive example is provided by H2AX which is by far the most abundant H2A variant in chromatin and is phosphorylated at a unique C-terminal motif in response to DNA damage (for review, see [19]). The C-terminal residues that distinguish the H2AX variant are not mutated at high frequency, and the number and pattern of reported missense mutations in H2AX is broadly similar to that of canonical H2A genes across most of the protein where they are identical. This suggests that, if H2AX acts as an oncohistone, it does so in a similar way to canonical H2A. It is surprising that so few missense mutations in the SQE motif of H2AX have been identified since deletions of *H2afx* predispose mice to tumour formation [20,21]. However, DNA damage responses are frequently misregulated or defective in cancer cells and these altered pathways can lead to increased dependency on the remaining DNA damage response pathways [22]. The lack of missense mutations in the SQE motif could therefore indicate that cancer cells become reliant on H2AX phosphorylation for their survival. This

illustrates the conundrum of oncohistones that only a narrow range of mutations may be tolerated yet able to drive cancer.

Mechanisms of oncohistone action

The improved annotation of histone genes makes it possible to map mutations to the genes encoding different protein isoforms, and therefore to assess whether isoforms with identical protein sequence have distinct genetic profiles. This reveals that amongst genes encoding the single H3.1 protein sequence, H3C2 is mutated with high frequency at E106, H3C10 is mutated most frequently at E98, and most H3C4 mutations are at R26 (Figure 2). Furthermore, these gene isoforms are mutated at different rates in tumours of different tissues.

The frequency of histone missense alleles is generally below 0.5, which is consistent with dominant genetics. However, since histone gene families are comprised of multiple isoforms distinguished by at most microheterogeneity in protein products, a dominant H3 allele needs to exert its effect in the context of over 29 non-mutated alleles.

A simple explanation for dominance in this context would be that oncogenic mutations are observed in the most highly expressed gene isoform. For example, the relatively highly mutated H2AC6 contributes ~20% of the H2A transcript in the cell lines surveyed (Figure 2). In contrast, H2BC5 is frequently mutated at positions not observed in other H2B gene isoforms yet contributes <5% of histone transcripts. Overall, the correlation between expression and mutation rate is weak across different gene isoforms (Figure 2). This indicates that the functional consequences of an oncohistone mutation must exert dominance over a large excess of wild type protein. Such a “super-dominant” behaviour constrains potential mechanisms by which oncohistones could exert their effects.

It is difficult to understand how a mutation affecting a few percent of the expressed protein could be incorporated into specific nucleosomes at the same locus in a large proportion of cells. This undermines the tempting interpretation that oncohistone-containing nucleosomes act locally to regulate access to discrete regulatory elements or processes.

A more attractive explanation for the super-dominant behaviour of oncohistones is that the presence of a low density of altered nucleosomes affects the function of an extended chromatin region. For example, a dominant effect would result by mutations inhibiting the action of a histone modifying enzyme and titrate the abundance of modifications globally as

seen for H3 K37M and PRC2 [15,16]. Another possibility is that alterations to the biophysical properties of chromatin could drive global effects on genome stability or transcription. For example, oncohistone mutations could affect the rate at which DNA and RNA polymerases transit through regions of chromatin, the efficiency of repair pathways, or the malleability of chromatin and its ability to adopt distinct three-dimensional conformations. Therefore, oncohistones linked to tissue-specific cancers would be a consequence of impaired function of a subset of tissue-specific enhancer elements that are dependent on these global chromatin features.

An important corollary is that data generated using experimental systems to explore oncohistone behaviour that use exogenous expression constructs (*e.g.* [13,23]) should be interpreted with this in mind, and experiments should aim to recapitulate physiological levels where possible.

Linking oncohistone expression and function

The influence of relative expression levels has an additional potential feature influencing their effects since expression levels could change during the evolution of cancer. H2AC1 is the most frequently mutated H2A-encoding gene, especially at R35 (Figure 2). None of the amino acids that are unique to H2AC1 are frequently mutated residues, and R35 was recognised as part of a three-dimensional cluster of mutated residues [11]. H2AC1 and H2BC1 are co-ordinately regulated and divergently transcribed genes whose expression is normally restricted to germ cells [24]. H2AC1 and H2BC1 incorporation generates to a more open chromatin structure [25] and expression of these gene isoforms can enhance stem cell reprogramming [24,26]. There is evidence that these isoforms are re-expressed in some cancer cells [27]. One attractive possibility, therefore, is that re-expression of these isoforms in somatic cells drives de-differentiation and the acquisition of stem cell-like properties to promote tumorigenesis, and this is exacerbated by H2AC1 R35 mutation. Expression of the mutant or wild-type isoforms could also be transient, with their effects exerted epigenetically in cancer cells even when expression is no longer evident.

Another route to alterations in histone expression in cancer cells is via gene amplification. Surprisingly, amplification of histone genes is more common than missense mutation. For example, H3C2 has 2 missense mutations but 26 amplifications amongst the 584 ovarian

cancer cases in the TCGA Pan Cancer study [28] and this trend appears to hold all four core histones, their variants and in linker histones. It might be expected that increasing histone gene copy number or deregulating their expression by changing context could provide a selective advantage in rapidly dividing tumour cells. Support for a possible causal role of histone gene amplification in genome instability comes from observations that defects in histone supply impede replication fork progression (for example, [29]). Amplifications of histone genes frequently span adjacent genes in the tight interspersed clusters which could multiply any effect on histone protein levels as well as complicate the differential expression of gene isoforms. How this overlays with missense mutations remains to be investigated.

Considerations and future perspectives

Oncohistones are mutations in histone encoding genes that create dominant drivers of cancer. Genomic analyses of histone mutations have led to a growing list of potential candidates. Cancers with very high tumour mutation burdens will harbour high numbers of passenger mutations, so not all missense mutations in canonical histones such as those collated in Figure 2 will be oncogenic drivers. Nevertheless, patterns are evident that support distinctive gene isoform specificity for the oncohistones. The frequently mutated residues are not identified as somatic mutations [30] suggesting they have functional significance. There is also a striking variation in the number of mutations for each histone type even though they have similar lengths and numbers of genes. H3 genes collectively having a much greater number of mutations that the other three core histone genes. Linker histone genes harbour even more mutations.

The mapping of mutations to individual histone gene isoforms also indicates the remarkable ability of small proportions of altered histones to exert effects in the presence of a substantial excess of wild type protein. Since it is not easy to imagine how such oncohistones could be strongly enriched in nucleosomes at specific loci, their super-dominant behaviour could reflect fine balances in global features of the chromatin substrate. This can provide a novel entry point to uncover functions of chromatin at a scale beyond individual nucleosomes.

The complexity of histone gene families including their microheterogeneity and tight interspersion in clusters has so far confounded attempts to assign functions to specific gene isoforms. Additional aspects such as frequency of amplifications within clusters and non-

polyadenylation of transcripts undermined simple transcriptome analyses. Recent progress means it is now clear that individual histone gene isoforms are genetically distinct, despite sometimes encoding identical proteins. A straightforward explanation for this paradox is that differential expression of histone gene isoforms defines their uniqueness. Systematic assessment of histone gene expression during oncogenesis that takes into account the rich properties of this fascinating gene family is likely to shed new light on how oncohistone mutations exert their effects.

Conflict of interest statement

Nothing declared.

Acknowledgements

The authors would like to thank Nicola Wiechens for analysis of transcriptome datasets. This work was supported by Cancer Research UK (C7905/A25715, JAD) and the Medical Research Council, UK (MR/SO21647/1, TOH).

References and recommended reading

Papers of particular interest have been highlighted as
* of special interest
** of outstanding interest

1. Alberts B: *Molecular biology of the cell* edn 4th. New York: Garland Science; 2002.
2. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ: **The human and mouse replication-dependent histone genes**. *Genomics* 2002, **80**:487-498.
3. Singh R, Bassett E, Chakravarti A, Parthun MR: **Replication-dependent histone isoforms: a new source of complexity in chromatin structure and function**. *Nucleic Acids Res* 2018, **46**:8665-8678.
4. Pinto DM, Flaus A: **The human canonical core histone catalogue**. BioRxiv doi:10.1007/978-90-481-3471-7_4
*This is a useful resource of human histone gene information and nomenclature.
5. Milo R, Jorgensen P, Moran U, Weber G, Springer M: **BioNumbers--the database of key numbers in molecular and cell biology**. *Nucleic Acids Res* 2010, **38**:D750-753.
6. Marzluff WF, Koreski KP: **Birth and Death of Histone mRNAs**. *Trends Genet* 2017, **33**:745-759.
7. Mendiratta S, Gatto A, Almouzni G: **Histone supply: Multitiered regulation ensures chromatin dynamics throughout the cell cycle**. *J Cell Biol* 2019, **218**:39-54.
8. Gunesdogan U, Jackle H, Herzig A: **Histone supply regulates S phase timing and cell cycle progression**. *Elife* 2014, **3**:e02443.
9. Mejlvang J, Feng Y, Alabert C, Neelsen KJ, Jasencakova Z, Zhao X, Lees M, Sandelin A, Pasero P, Lopes M, et al.: **New histone supply regulates replication fork speed and PCNA unloading**. *J Cell Biol* 2014, **204**:29-43.
10. Miles DM, Desdouets C, Geli V: **Histone stress: an unexplored source of chromosomal instability in cancer?** *Curr Genet* 2019, **65**:1081-1088.
11. Nacev BA, Feng L, Bagert JD, Lemiesz AE, Gao J, Soshnev AA, Kundra R, Schultz N, Muir TW, Allis CD: **The expanding landscape of 'oncohistone' mutations in human cancers**. *Nature* 2019, **567**:473-478.
**This paper provides an analysis of histone protein mutations that uncovered the scope and range of oncohistone mutations and their patterns in cancer.
12. Funato K, Tabar V: **Histone Mutations in Cancer**. *Annual Review of Cancer Biology* 2018, **2**:337-351.

13. Bennett RL, Bele A, Small EC, Will CM, Nabet B, Oyer JA, Huang X, Ghosh RP, Grzybowski AT, Yu T, et al.: **A Mutation in Histone H2B Represents a New Class of Oncogenic Driver**. *Cancer Discov* 2019, **9**:1438-1451.
**This paper performed an analysis of cancer samples and identified potential oncohistone mutations. The authors characterized the H2B-E76K mutation and provide evidence of nucleosome dysfunction and oncogenic effects on gene expression.
14. Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, Qu C, Ding L, Huether R, Parker M, et al.: **Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas**. *Nat Genet* 2012, **44**:251-253.
15. Chan KM, Fang D, Gan H, Hashizume R, Yu C, Schroeder M, Gupta N, Mueller S, James CD, Jenkins R, et al.: **The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression**. *Genes Dev* 2013, **27**:985-990.
16. Lewis PW, Muller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ, Allis CD: **Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma**. *Science* 2013, **340**:857-861.
17. Zhao Z, Shilatifard A: **Epigenetic modifications of histones in cancer**. *Genome Biol* 2019, **20**:245.
18. Arimura Y, Ikura M, Fujita R, Noda M, Kobayashi W, Horikoshi N, Sun J, Shi L, Kusakabe M, Harata M, et al.: **Cancer-associated mutations of histones H2B, H3.1 and H2A.Z.1 affect the structure and stability of the nucleosome**. *Nucleic Acids Res* 2018, **46**:10007-10018.
19. Pinto DM, Flaus A: **Structure and function of histone H2AX**. *Subcell Biochem* 2010, **50**:55-78.
20. Bassing CH, Suh H, Ferguson DO, Chua KF, Manis J, Eckersdorff M, Gleason M, Bronson R, Lee C, Alt FW: **Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors**. *Cell* 2003, **114**:359-370.
21. Celeste A, Difilippantonio S, Difilippantonio MJ, Fernandez-Capetillo O, Pilch DR, Sedelnikova OA, Eckhaus M, Ried T, Bonner WM, Nussenzweig A: **H2AX haploinsufficiency modifies genomic stability and tumor susceptibility**. *Cell* 2003, **114**:371-383.
22. O'Connor MJ: **Targeting the DNA Damage Response in Cancer**. *Mol Cell* 2015, **60**:547-560.
23. Pathania M, De Jay N, Maestro N, Harutyunyan AS, Nitarska J, Pahlavan P, Henderson S, Mikael LG, Richard-Londt A, Zhang Y, et al.: **H3.3(K27M) Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas**. *Cancer Cell* 2017, **32**:684-700 e689.
24. Shinagawa T, Takagi T, Tsukamoto D, Tomaru C, Huynh LM, Sivaraman P, Kumarevel T, Inoue K, Nakato R, Katou Y, et al.: **Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells**. *Cell Stem Cell* 2014, **14**:217-227.
25. Padavattan S, Shinagawa T, Hasegawa K, Kumasaka T, Ishii S, Kumarevel T: **Structural and functional analyses of nucleosome complexes with mouse histone variants TH2a and TH2b, involved in reprogramming**. *Biochem Biophys Res Commun* 2015, **464**:929-935.
26. Huynh LM, Shinagawa T, Ishii S: **Two Histone Variants TH2A and TH2B Enhance Human Induced Pluripotent Stem Cell Generation**. *Stem Cells Dev* 2016, **25**:251-258.

27. Wang T, Chuffart F, Bourova-Flin E, Wang J, Mi J, Rousseaux S, Khochbin S: **Histone variants: critical determinants in tumour heterogeneity**. *Front Med* 2019, **13**:289-297.
28. Consortium ITP-CAoWG: **Pan-cancer analysis of whole genomes**. *Nature* 2020, **578**:82-93.
29. Prado F, Maya D: **Regulation of Replication Fork Advance and Stability by Nucleosome Assembly**. *Genes (Basel)* 2017, **8**.
30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al.: **The mutational constraint spectrum quantified from variation in 141,456 humans**. *Nature* 2020, **581**:434-443.
31. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL: **Genomewide characterization of non-polyadenylated RNAs**. *Genome Biol* 2011, **12**:R16.
32. Fritz AJ, Gillis NE, Gerrard DL, Rodriguez PD, Hong D, Rose JT, Ghule PN, Bolf EL, Gordon JA, Tye CE, et al.: **Higher order genomic organization and epigenetic control maintain cellular identity and prevent breast cancer**. *Genes Chromosomes Cancer* 2019, **58**:484-499.
33. Michel M, Demel C, Zacher B, Schwalb B, Krebs S, Blum H, Gagneur J, Cramer P: **TT-seq captures enhancer landscapes immediately after T-cell stimulation**. *Mol Syst Biol* 2017, **13**:920.
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics* 2013, **29**:15-21.
35. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al.: **Ensembl 2020**. *Nucleic Acids Res* 2020, **48**:D682-D688.
36. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al.: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data**. *Cancer Discov* 2012, **2**:401-404.

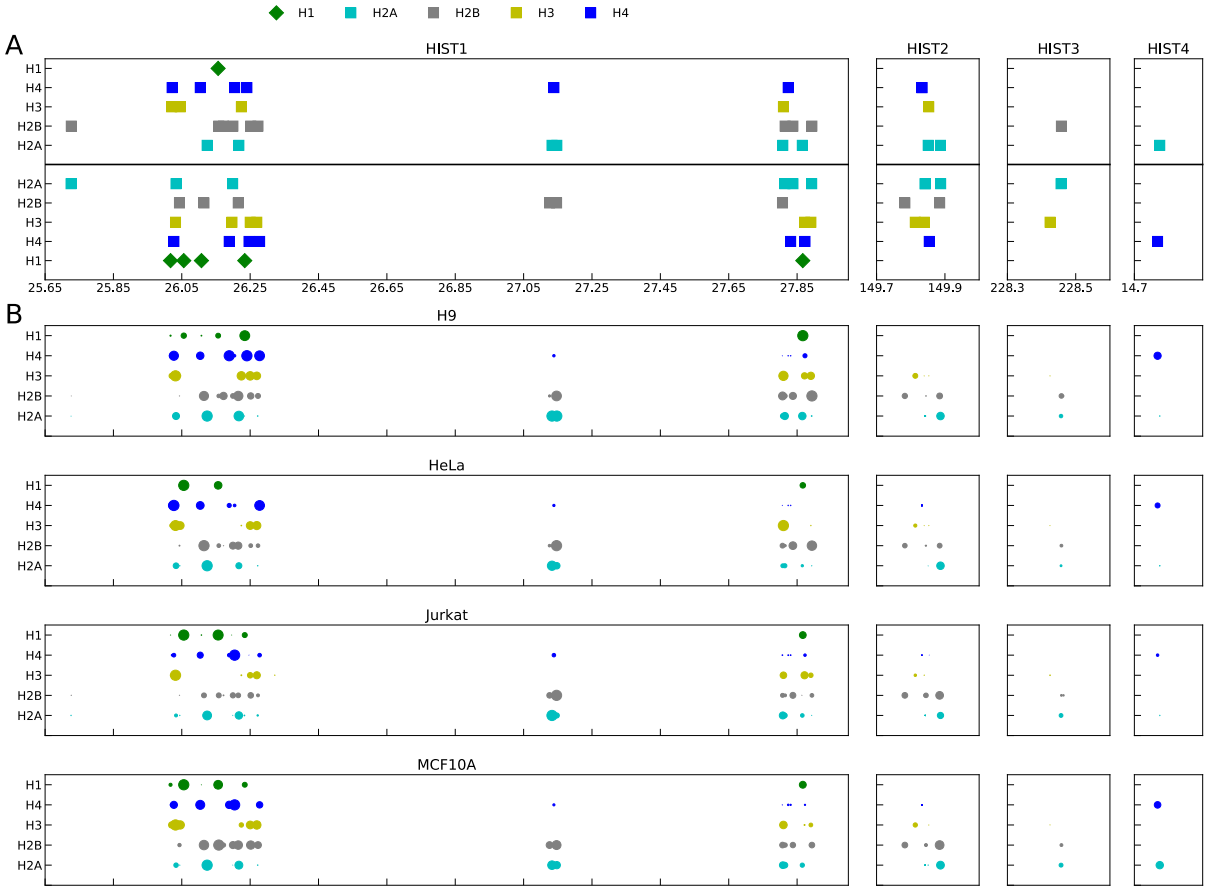

Figure 1. Location and expression of canonical histone genes. A. Location of human canonical genes in histone loci HIST1-4 on chromosomes 6, 1, 1 and 12 respectively in genome release GRCh38.p13. B. Relative expression of genes for each type indicated by circle area using mapped read counts in total RNA-seq of H9 and HeLa [31] and of MCF10A [32] and TT-seq of Jurkat [33] from STAR [34] with ENSEMBL release 99 [35].

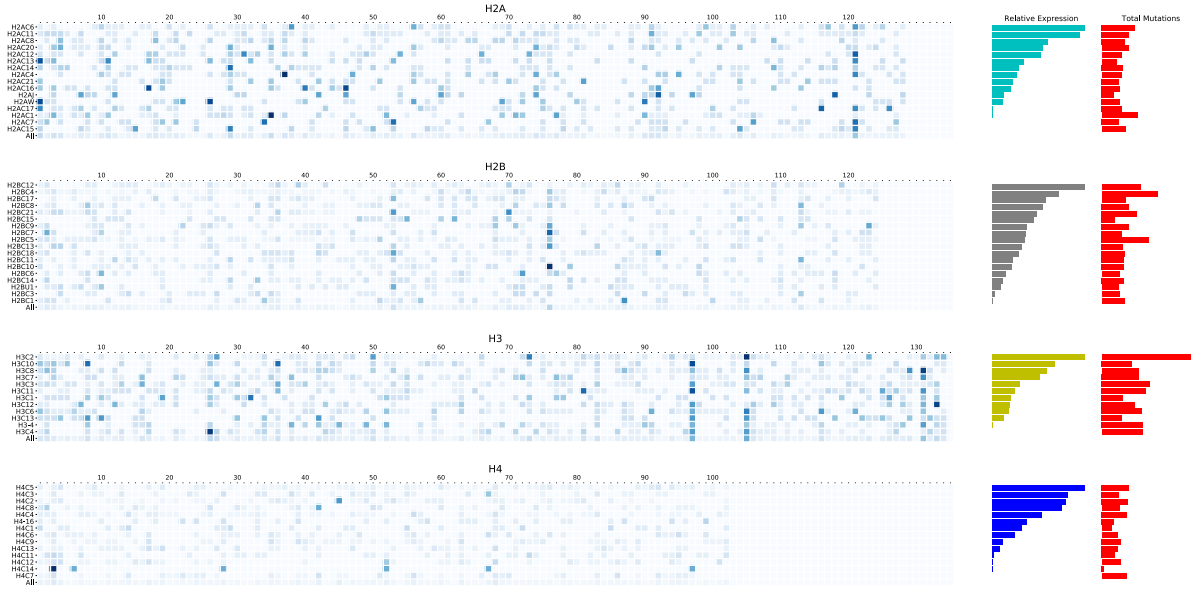

Figure 2. Mutated residues in canonical core histone isoforms. Heat map of relative count of missense mutations within each isoform residue for 4800 unique case mutations from cBioPortal [36] in September 2020 excluding cell lines. Isoforms are sorted by decreasing mean gene expression levels across 4 cell lines in central bar plot coloured as in Figure 1B. Count of total mutations for each isoform is in right bar plot, with relative mutations per residue aggregated from all isoforms of the histone type in the bottom row of the heat map.