The expanding landscape of 'oncohistone' mutations in human cancers

Benjamin A. Nacev[1,2], **Lijuan Feng**[2,*], **John D. Bagert**[3,*], **Agata E. Lemiesz**[2,†], **JianJiong Gao**[1,†], **Alexey Soshnev**[2], **Ritika Kundra**[1], **Nikolaus Schultz**[1], **Tom W. Muir**[3,††], and **C. David Allis**[1,2,††]

[1]Memorial Sloan Kettering Cancer Center, New York, NY, 10065 USA.

[2]Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY 10065 USA.

[3]Department of Chemistry, Princeton University, Princeton, NJ 08544 USA

Abstract

Mutations in epigenetic pathways are common oncogenic drivers. Histones, the fundamental substrate for chromatin-modifying and remodeling enzymes, are mutated in tumors including in gliomas, sarcomas, head and neck cancers, and carcinosarcomas. Classical 'oncohistone' mutations occur in the N-terminal tail of histone H3 and impact the function of Polycomb Repressor Complexes 1 and 2. However, the prevalence and function of histone mutations in additional tumor contexts is unknown. Here we show that somatic histone mutations conservatively occur in ~ 4% of tumors of diverse types and in critical regions of histone proteins. Mutations occur in all four core histones, in both the N-terminal tails and globular histone fold domains, and at or near residues that harbor important post-translational modifications. Many globular domain mutations are either homologous to yeast mutants that abrogate the need for SWI/SNF function, occur in the key regulatory 'acidic patch' of histone H2A and H2B, or are predicted to disrupt the H2B-H4 interface. The histone mutation dataset (https://bit.ly/2GXH5Ve) and the hypotheses presented herein on the impact of the mutations on important chromatin functions should serve as a resource and starting point for the chromatin and cancer biology fields in exploring an expanding role of histone mutations in cancer.

The fundamental repeating subunit of chromatin is the nucleosome, a histone octamer which is wrapped by 147 base pairs of DNA.[1] The density and positioning of nucleosomes sterically determine the ability of cellular machinery access to the genome (Figure 1a). Consequently, chromatin structure plays a critical role in diverse processes including activating or repressing transcription to control functions such as cell fate, the cell cycle, and DNA damage repair[2].

A critical component of chromatin-mediated regulation utilizes histone post-translational modifications (PTMs), in which histones integrate cellular signals to choreograph chromatin-dependent functions[3,4]. Given the regulatory role of chromatin for all DNA-templated processes, it is not surprising that the protein machinery that 'writes', 'reads', and 'erases' these histone marks is frequently altered in cancer, and in many cases these mutations are oncogenic drivers or contributors to tumor progression[5]. Mutations in the histones themselves have also recently been linked to cancers, namely the discovery that mutations in histone H3 occur with high genetic penetrance within rare pediatric gliomas and sarcomas[6–8](Figure 1b).

These mutations, some of which act in a dominant fashion, have been deemed 'oncohistones' and include H3K27M, which was identified in 78% of diffuse intrinsic pontine gliomas, as well as H3G34V/R, which also occurs in pediatric glioblastomas[6,8]. Among sarcomas, the histone H3 variant H3.3 is mutated at lysine 36 (H3.3K36M) in 95% of chondroblastomas and at glycine 34 (H3.3G34W/L) in 92% of giant cell tumors of the bone[7]. Oncohistones have also been observed in diffuse large B-cell lymphomas (histone H1), head and neck cancers (H3K36M), and in carcinosarcomas (H2A and H2B)[9–11]. A striking feature of these founding oncohistone mutations is their location at or near key regulatory PTMs in the histone tails, suggesting they might disrupt the 'reading', 'writing', and/or 'erasing' of these marks[12]. Work done in several laboratories, including ours, demonstrates that the H3K27M mutation found in gliomas acts as a dominant negative inhibitor of the EZH2 subunit of the Polycomb Repressor Complex 2 (PRC2) 'writer' leading to a loss of transcriptional silencing through a global reduction in H3K27 tri-methylation (H3K27me3)[12,13].

The H3K36M mutation leads to global loss of H3K36 di- and tri-methylation and also increases H3K27me3 levels, which at intergenic regions promotes the recruitment of the H3K27me3 reader complex, Polycomb Repressor Complex 1 (PRC1), away from gene-associated H3K27me3[14,15]. This aberrant recruitment leads to de-repression of Polycomb-regulated genes, which blocks mesenchymal differentiation, and is sufficient to promote a sarcoma-like tumor in a mouse xenograft[14]. Notably, the H3K36M oncogenic driver mutation occurs at a high frequency in chondroblastomas (a low tumor mutation burden tumor), but only rarely in head and neck squamous cell carcinomas (a high tumor mutation burden tumor)[7,10,16]. Thus, the tumor type frequency of a histone mutation is not necessarily a clear predictor for biological importance.

The emerging oncohistone field raises the question of whether histone mutations exist in other cancers, and if so, whether those mutations are confined to the histone tails at or near

known PTM sites, as in the current paradigm. Thus, as a hypothesis-generating effort, we sought to catalogue and characterize the landscape of missense histone mutations across multiple cancer types. The mutation data reported herein are available through cBioPortal (https://bit.ly/2GXH5Ve) as an interactive interface in addition to Supplementary Tables. We analyzed this dataset to propose hypotheses and potential mechanisms underlying the role of these oncohistones in altering chromatin structure and function, potentially contributing to the development of the tumors in which they are observed.

Analysis of both publicly available tumor sequencing databases and previously unreported data from the Memorial Sloan Kettering internal sequencing effort (MSK-IMPACT) revealed a total of 4205 histone missense mutations in 3143 samples from 3074 unique patients across 183 specific tumor types (Supplementary Tables 1 and 2). Histone mutations were identified in all core histone families and 598 samples harbor multiple histone mutations (Extended Data Figure 1).

In total, 1039 unique patients were identified in the MSK-IMPACT clinical sequencing cohort, with the remainder found in 98 published studies available through cBioPortal (Supplementary Table 2). In the MSK-IMPACT database, which we note currently includes less than one third of histone genes, histone missense mutations are nonetheless identified in 3.8% of tumor samples. This is approximately the same prevalence as somatic mutations in a number of cancer-associated genes in the same cohort including BRCA2, TET2, SMAD4, and NOTCH1. Because histone mutations are distributed over dozens of histone genes, we suggest that they might lack visibility in rank-ordered lists of individual genes often reported in tumor sequencing studies. This may contribute to the relative lack of emphasis on histone mutations in past studies despite a notably high mutation rate.

Re-identification of known oncohistones as well as novel lysine mutations patterns

To generate hypotheses about the functional role of these histone mutations, we conducted subsequent analysis on samples with tumor mutation burden (TMB) of ≤10 mutations/Mb (Supplementary Table 3) (unless otherwise specified) in order to mitigate confounding effects from highly mutated tumors. Above a 10 mutations/Mb threshold, the number of additional captured mutations at H3K27, H3G34, and H3K36 decreases dramatically, which supports the use of this specific cutoff (Extended Data Figure 2). This refined dataset includes 1921 tumor-associated histone mutations, many of which are at relevant tumor allele frequencies and occur across both common and rare tumor types (Extended Data Figure 1).

Importantly, this analysis re-identified known oncohistone mutations including H3K27M and H3G34R/V in gliomas, H3G34W in osteosarcoma, and H3K36M in head and neck cancers (Figure 2a and Supplementary Table 3). Interestingly, within H3 there are a number of residues including H3E105, H3E97, and H3R26, which are mutated at rates similar to the known oncohistones (Figure 2b). We also observed the original set of oncohistones in tumor types where they have not been previously appreciated. This includes H3K27M in melanoma and acute myeloid leukemia, H3G34V in ovarian cancer, and H3K36M in

melanoma, bladder, and colorectal cancer. Thus, known driver mutations in histones, which occur at a high frequency in rare cancers, also exist at a lower frequency in more common cancers.

In keeping with the pattern of lysine mutations previously observed in oncohistones, we note similar mutations at H3K4: out of a total of 9 mutations at this site, 8 were a H3K4M/I substitution. One K-to-M/I mutation was also observed at H3K18 and at H4K12 raising the possibility that the functional effects associated with known K-to-M/I changes (i.e. creation of a substrate-derived methyltransferase inhibitor) may extend to additional contexts[12]. Beyond the K-to-M/I paradigm, we also observe new lysine mutation patterns. For instance, in H2A, 10 of 12 mutations at K74 and K75 are lysine-to-asparagine mutations. Whether the frequency of K-to-N mutagenesis indicates a functional significance akin to K-to-M mutations remains to be determined.

Histone mutations occur at N-terminal as well as globular domain residues

Similar to the better studied oncohistones, many of the more prevalent point mutations within any given core histone family involve the N-terminal tail domains and occur at or near the site of known PTMs (Figure 2a, Extended Data Figures 7 and 9). For instance, H3R26 (19 mutations) is among the most commonly mutated residues in our dataset and is adjacent to H3K27 (18 mutations) (Table 1). This implies that these N-terminal tail mutations may influence the function of the chromatin 'writing', 'reading', or 'erasing' machinery that operates at or nearby those residues, which is consistent with the paradigm established by H3K27M and H3K36M.

Interestingly, 4 out of 5 of the most commonly mutated residues are in a globular domain: H3E105 (26 mutations), H2BE76 (24 mutations), H3E97 (22 mutations), and H2BE113 (21 mutations) (Table 1, Extended Data Figure 7). A similar trend is observed within each core histone family (Figure 2b). Across all tumor types, mutations in the globular domain of H3 are observed more frequently than those in amino-terminal tail, which include the well-characterized K27M and K36M mutations. This trend is not limited to H3. In H4, for example, the highest frequency mutations, R3C, L49F, S1C, and K79N involve both the N-terminal tail and its globular domain (Figure 2b). This raises the possibility that histone mutations in tumors may have effects beyond perturbation of PTM-associated functions.

Trends towards tumor type specific mutation signatures

To explore the pattern of histone mutations between cancer types, we performed unsupervised hierarchical clustering of our dataset without using TMB cutoffs as this could bias the analysis against tumor types with generally higher mutation rates (Extended Data Figures 3–7). As expected, in many cases mutations at specific residues occur independently of cancer type. However, there are also instances of differential clustering. For instance, the mutational pattern of H3 residues in pancreatic cancer appears quite different from that of cervical cancer (Extended Data Figure 3). Conversely, we also define clusters of mutations between cancer types, including shared H2BE76 mutations in bladder and cervical cancer and H2AE121 mutations in bladder, breast, and head and neck cancer (Extended Data

Figures 5 and 6). Further, there is a distinct cluster of four different hematologic malignancies defined by mutations at H3S86 (Extended Data Figure 3). Taken together, these observations raise the intriguing possibility of a rare lineage preference for specific point mutations within histone families.

Sin⁻ mutations

Classic studies in yeast identified histone H3 and H4 mutants that abrogate the need for the SWI/SNF remodeling complex in regulating gene expression[17]. These so-called *Sin⁻* (SWI/SNF Independent) mutations cluster in the globular domains of histone H3 (residues 105 to 118) and histone H4 (residues 43 to 45) (Figure 3a). One of the *Sin⁻* mutations, H3E105K (conserved from yeast to humans), occurs at the most commonly mutated residue in the dataset and H3E105K/Q has been annotated in cBioPortal as a three-dimensional hotspot mutation based in part on a computational method for modeling mutations on structural data[18]. Additional yeast H3 *Sin⁻* mutations occur at H3R116 and H3T118, which are also mutated in tumors. With regard to histone H4, there are six mutations at H4R45, including H4R45C, which is a known yeast *Sin⁻* mutant[17]. Notably, H4R45C-containing nucleosomes have been shown to abolish higher order chromatin folding[19]. Thus, the finding of tumor mutations corresponding to the yeast *Sin⁻* mutants raises the hypothesis that they may disrupt chromatin folding and DNA packaging into nucleosomes. Given the well-established role of the SWI/SNF complex in regulating gene expression, these mutations may also play a role in oncogenesis through aberrant expression of oncogenes and/or disrupting developmental programs[20]. Along this line, we note that subunits of mammalian SWI/SNF subunits are mutated at relatively high frequency in a large number of human cancers[21]. Whether *Sin⁻* patch oncohistone mutations act to enhance or suppress SWI/SNF subunit mutations in these cancers remains to be determined.

Acidic Patch mutations

We observed frequent mutations within a nucleosome anchoring point commonly referred to as the "acidic patch" formed by six H2A and two H2B residues (Figure 3b and Extended Data Figure 7). Acidic patch mutations at residues H2BE113, H2AE92, and H2AE56, occur at high frequencies for their respective histone families and H2BE113 is the fifth most common site of mutations in the dataset. Notably, H2AE56K/Q mutations were previously reported in human uterine and ovarian carcinosarcomas and are now reported in our dataset in non-small cell lung caner, renal cell carcinoma, small cell lung cancer, head and neck cancer, pancreas, and rectal cancer[11].

Given the important function of the negatively charged acidic patch surface, tumor-associated acidic patch mutations are hypothesized to affect multiple essential biological processes including chromatin condensation and folding, nucleosome remodeling, cell division, transcriptional silencing, and DNA damage repair[22]. In fact several of the residues mutated in our dataset, including H2AE61, H2AD90, and H2AE92, were shown to impair chromatin remodeling by ISWI (Figure 2a)[23]. Since acidic-patch-dependent processes are critical for cell identity, differentiation, and genomic and epigenomic integrity, acidic patch mutations may promote oncogenesis by disruption of one or more of those pathways.

Mutations that potentially alter nucleosome structure

Mutations within the globular core domain of the histones have the potential to impact three-dimensional structure. With this in mind, we created a proximity plot based on the nucleosome structure and an enriched set of histone mutations with a residue-specific mutation count ≥2.5-fold the median count/residue in each histone family (Extended Data Figure 8, Figure 3c). Remarkably, many of the most highly mutated residues, although quite distant in primary amino sequence, are close in space within the folded nucleosome structure. This raises the intriguing possibility that multiple mutations at distinct histone sites may functionally converge, either to perturb the functionality of PTMs in the immediate vicinity (Extended Data Figure 9) or to directly disrupt the structure of the nucleosome (Extended Data Figure 10). H2BE76, for example, lies at the tetramer-dimer interface between H2B and H4 and engages with residues R92 and D68 of H4 in a hydrogen-bonding network that would be abolished by H2BE76K/Q mutations (Figure 3d). Indeed, very recent work has shown that the H2BE76K mutation destabilizes nucleosomes and perturbs the local structural arrangement of H4R92[24], supporting our hypothesis that mutations at H4R92 and H4D68 may have similar functional effects. Notably, H2BE76 is the most highly mutated residue in the H2B family and H4D68 and H4R92 are the second most and fifth most mutated residue in the H4 family, respectively (Extended Data Figure 7). In addition, a nearby salt bridge between H2BE71 and H4K91 at the same interface is disrupted by acetylation and ubiquitylation of H4K91, affecting chromatin assembly and DNA damage repair[25–27]. Interestingly, germline mutations at H4K91 (H4K91R/Q) were recently found to cause a severe developmental syndrome[28]. In our dataset H2BE71 is the fifth most mutated residue in H2B, demonstrating a striking convergence of cancer mutation frequency onto this portion of the H2B-H4 interface.

Continuing with the theme of structural perturbations, we note there are a number of cases in which a glycine or proline residue is introduced into the globular domain of the histone, for example at H2AR29 and H4R39. Alterations of this type are expected to disrupt the alpha-helical secondary structure of the histone fold.

Candidate neomorphs

Another notable pattern is that glutamic acid residues are the second most mutated in our study and are frequently substituted to lysine (47 % of cases) or glutamine (34 % of cases) (Figure 2a and Extended Data Figure 7). We hypothesize that these mutations may function as neomorphs in which the ectopic lysine could be post-translationally modified, or in which the glutamine functions as an acetyl-mimic to aberrantly recruit reader complexes (Figure 3e). That said, we recognize that the mutation patterns reported here may well be influenced by specific mutagens, nucleotide composition, codon biases, chromosomal position, and tissue type. Nonetheless, we look forward to experimental testing of this hypothesis.

Mutations identified by 'a posteriori' approaches

While some candidates for histone driver mutations may be hypothesized *a priori*, as above, there may also be less intuitively apparent yet functionally important mutations within the

dataset. Given that tumors typically harbor two to eight driver mutations, focusing on histone mutations occurring in a low mutational background may be useful in refining the list of candidate histone driver mutations[29]. Thus, we analyzed the subset with a TMB ≤2 mutations/Mb (Supplementary Table 4). Notably, among the histone mutations seen in the context of a low TMB are the known H3K27M and H3G34W oncohistones (though not H3K36M) suggesting that a low TMB cutoff captures a subset of functionally relevant histone mutations. Other mutations in this subset include H3E105K/Q, mutations at H3 N-terminal residues at or near PTM sites including R2, R8, K18, and R26, as well as residues in the acidic patch such as H2A E56, E64, E91, and E92 and H2B E105 and E133. Notably, this low TMB threshold list also includes mutations without a clear functional correlation raising questions regarding as yet unrecognized function. These types of mutations are likely to serve as an excellent starting point for future mechanistic investigations.

As an additional methodology to identify mutations that are potentially functional, we applied a previously published computational methodology for identifying three dimensional (3D) clusters of mutations[18]. This approach identified regions of all four core histones, which harbor potential functional mutations, especially those that occur at lower frequency (Table 2). These regions include the portions of the H3 *Sin⁻* patch and PTM-rich segments of the H3 and H4 tails, which bolsters the argument that mutations at or near the sites of regulatory PTMs may have functional significance. Importantly, this analysis also highlights stretches of the H2A and H2B globular domains that would not have been recognized otherwise but are significantly mutated. We suggest that mutations occurring in those regions warrant further investigation into functional consequences.

'Missing' mutations

Finally, some residues are notably mutationally silent in the dataset even taking into account tumors with high TMB (Supplementary Table 1). For example, in H2A there are no mutations at F25, A45, L85, V107, L115, and T120 (exclusive of variant sequences at these positions). The relative underrepresentation of mutations at these positions leads us to hypothesize that these residues, and similarly underrepresented residues in other core histones, may harbor important functions that are lost when mutated, thereby causing a fitness disadvantage during tumor development. This implies that there may also be dependency of these tumor cells on the machinery that writes, erases, and reads the post-translational marks at or near these residues. Targeting this machinery with small molecules could represent a previously unrecognized therapeutic opportunity.

Considering histone mutations in the context of chromatin

We have previously hypothesized that the combinatorial nature of histone PTMs and the physiologic context of histones within a nucleosomal polymer invoke unique biological consequences[3,30]. Thus, we favor the view that histone mutations can perturb important chromatin-mediated processes, even at low concentrations compared to normal histones. For instance, incorporation of a defective polymer subunit (i.e. a histone with a single amino acid substitution) acting as a nucleation site, could lead to amplified effects on chromatin due to both the biophysical and functional interdependency of nucleosomal subunits, the

latter of which is attributed to the histone code[3]. Alternatively, a mutant histone may behave as an aberrant boundary element that disrupts the activity of a PTM writer, thereby impairing (either directly or indirectly) the propagation of histone marks (Figure 4). In support, the H3K36M oncohistone causes a redistribution of PRC1 by altering the balance of H3K27me3 between intragenic and genic regions[14].

We suggest that histone mutants – by virtue of incorporation into chromatin – are primed to have dominant negative effects on chromatin-dependent processes. This is in contrast to a hypothetical mutation in one of a dozen redundant genes for a signaling kinase, which would likely be compensated by independent functional gene products. Thus, interpreting histone missense mutations using common paradigms of cancer genetics does not necessarily account for the chromatin polymer into which mutant histone proteins are incorporated. We look forward to the development of new models for histone mutations in this context, which we hope to facilitate with the work we present herein.

In summary, these analyses highlight a significantly greater number of histone mutations in human tumors than had previously been recognized. Given the key role that histones play in cell fate and differentiation as the substrates for epigenetic machinery, histone mutations may cause the dysregulation of these processes, a step central to oncogenesis[31]. While many of the mutations in the dataset may ultimately be passengers, the nature of a subset of the mutations we analyzed has enabled us to generate testable hypotheses about possible functional roles of these oncohistones in tumorigenesis. Understanding mechanisms underlying a subset of these mutations may reveal how disrupting histone/nucleosomal function may contribute to oncogenic transformation.

Methods:

Data on missense mutations in human histone genes were compiled from publicly available tumor sequencing databases on cBioPortal.org[32,33] as well as unreported data from MSK-IMPACT[34,35] by querying the Memorial Sloan Kettering private instance of cBioPortal as of July 18th 2018. The query gene list was formulated based on histone genes listed in the UniProt database (Supplementary Table 5). The prevalence of histone missense mutations in the MSK-IMPACT cohort was calculated based on data available via cBioPortal on July 16, 2018. For the MSK-IMPACT cohort, only de-identified data was collected and the publication of MSK-IMPACT data in this study has been approved by MSK-IMPACT data usage committee. Please see the previous reporting of the MSK-IMPACT resource for methodology and statements regarding human subject use, consent, and the institutional IRB protocol review for NCT01775072[34].

Tumor samples without sequenced matched normal were excluded. Except for the MSK-IMPACT samples, only whole-genome or whole-exome sequenced (WES) samples were included. Non-synonymous tumor mutational burden of a sample (mutations per million base pairs) was calculated by dividing the number of non-synonymous mutations by the sequenced base pairs (30Mbp for WES samples, and 0.98–1.19Mbp for MSK-IMPACT samples depending on the panel version; see Supplementary Table 1). Next, the data was processed to remove duplicate samples, standardize the cancer type nomenclature (see

Curated_Main_Cancer_Type in Supplementary Table 1), and renumber the amino acid residues based on the convention within the histone field of not including the initiator methionine in the residue count. For clarity of data presentation, histone H2A variants were analyzed only in total mutation calculations, in global allele frequency distribution plots, and global TMB distribution plots. Because in some cases, the sequencing of multiple tumor samples from the same patient occurred intentionally, presumably for either clinical or research indications, a per patient subset of the data was created representing the union of mutations across multiple samples for individual patients (Supplementary Tables 2–4) and used for analysis unless otherwise stated.

3D hotspot analysis and associated statistical tests was performed as previously described[18].

Extended Data
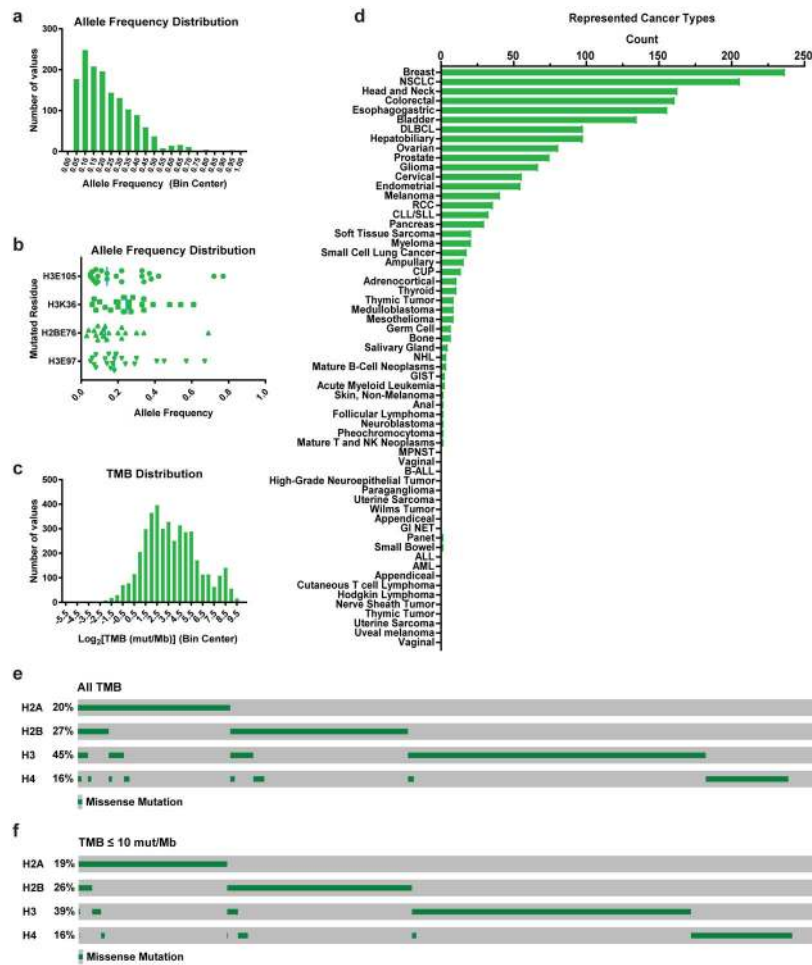
Extended Data Figure 1. Sample characteristics.

One sample per patient where TMB ≤10 mut/Mb is shown except (c) where all samples are represented regardless of TMB. a) Tumor allele frequency distribution 1452 of 1921 tumors where allele frequency is available based on publicly available data. b) Detailed tumor allele frequency distribution for the four most frequency mutated residues. Blue bars represent the median. c) TMB distribution. d) Tumor type distribution. For display purposes, the main cancer types are used in place of detailed cancer type. e) Oncoprint of the distribution of histone mutations between core families on a per patient level for all TMB and (f) for TMB ≤10 mut/Mb. For display purposes, H2A variants and H3.5 are not shown.
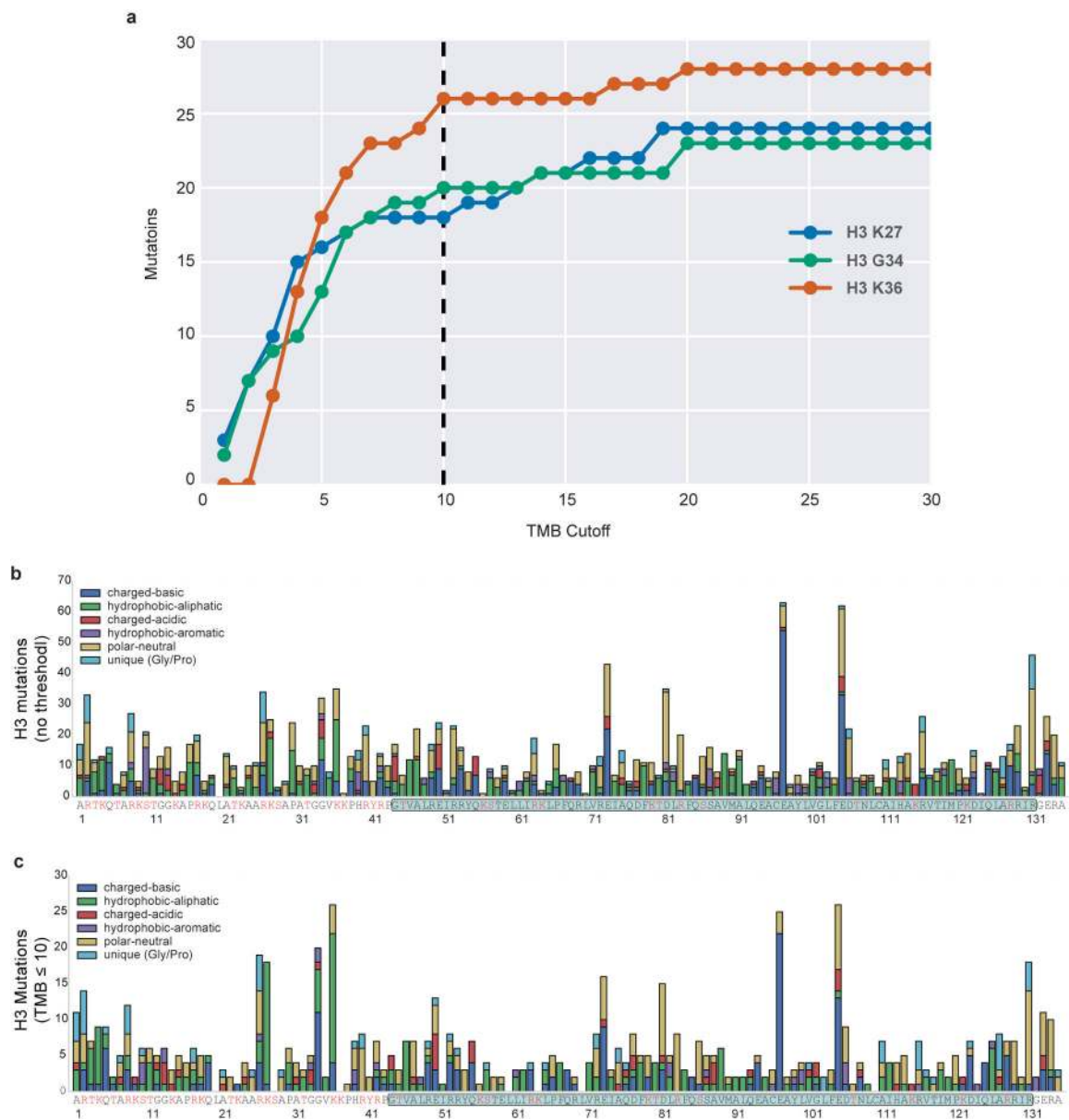
Extended Data Figure 2. Validation of TMB ≤10 as an analysis threshold.
a) For known oncohistones, the number of mutations captured reaches a plateau at TMB >
10 mut/Mb. b) Histogram plot of H3 mutation distribution on a per patient level without a
TMB threshold and (c) with a TMB ≤10 threshold shows enrichment of known
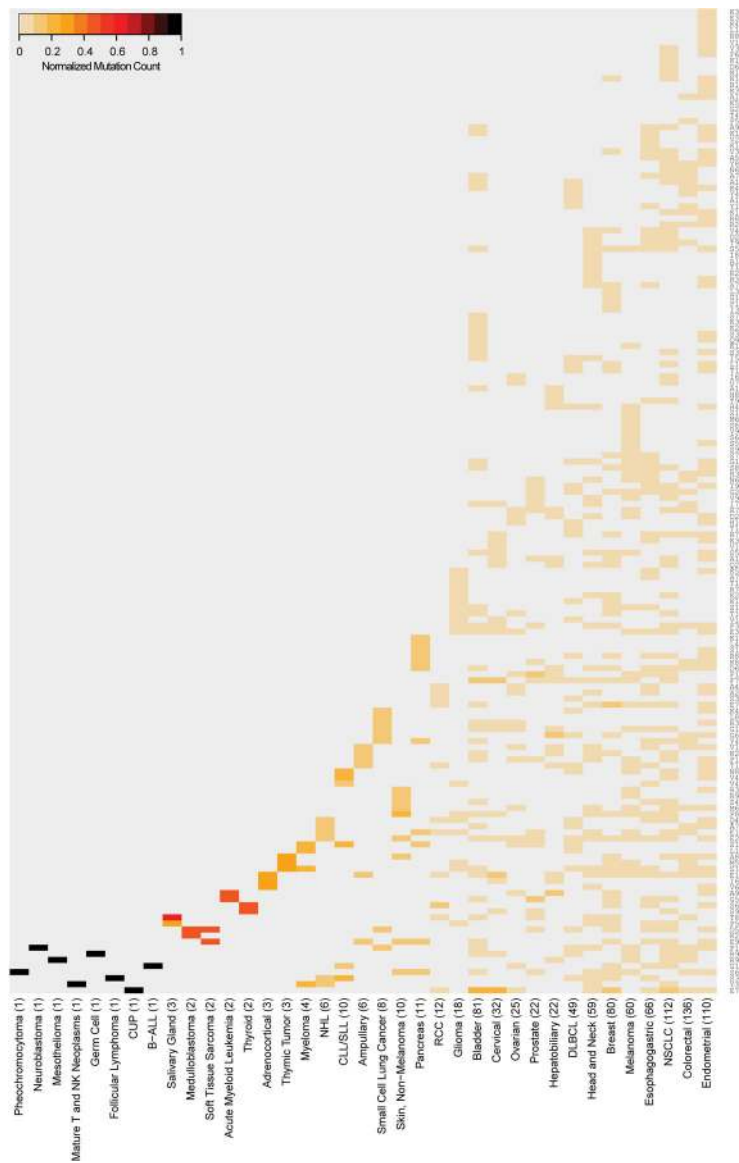oncohistones as well as additional mutations compared to background.
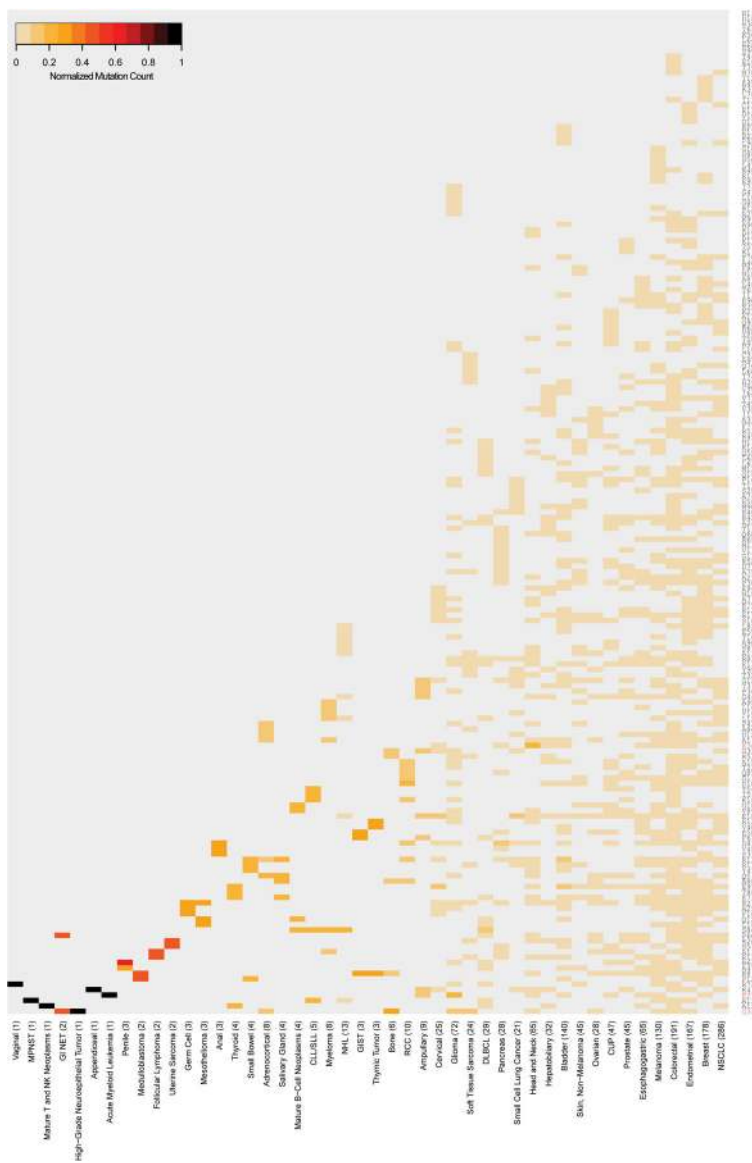
Extended Data Figure 3.
Heatmap of histone H3 mutations with individual residue labels. Color intensity indicates normalized mutation count (#mutations at residue/#samples per cancer type). Red labels indicate positions of known oncohistones. Per patient data with all TMB is plotted. The numbers of tumors sequenced is indicated following the tumor type label.
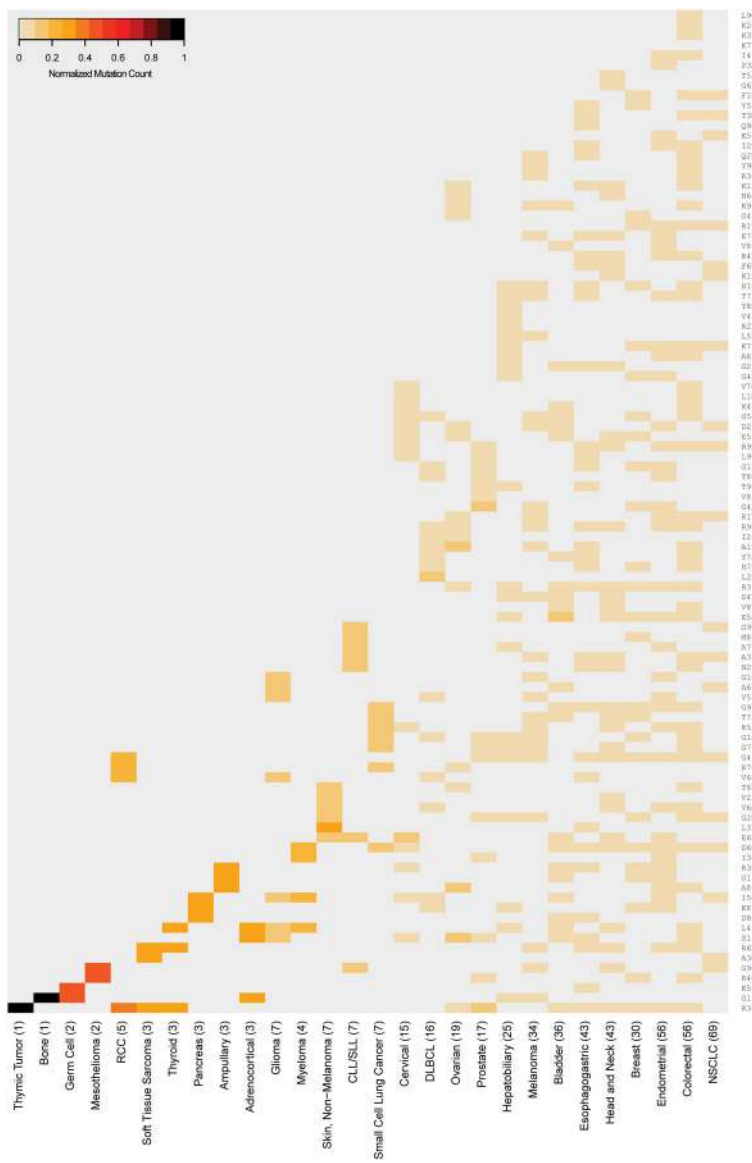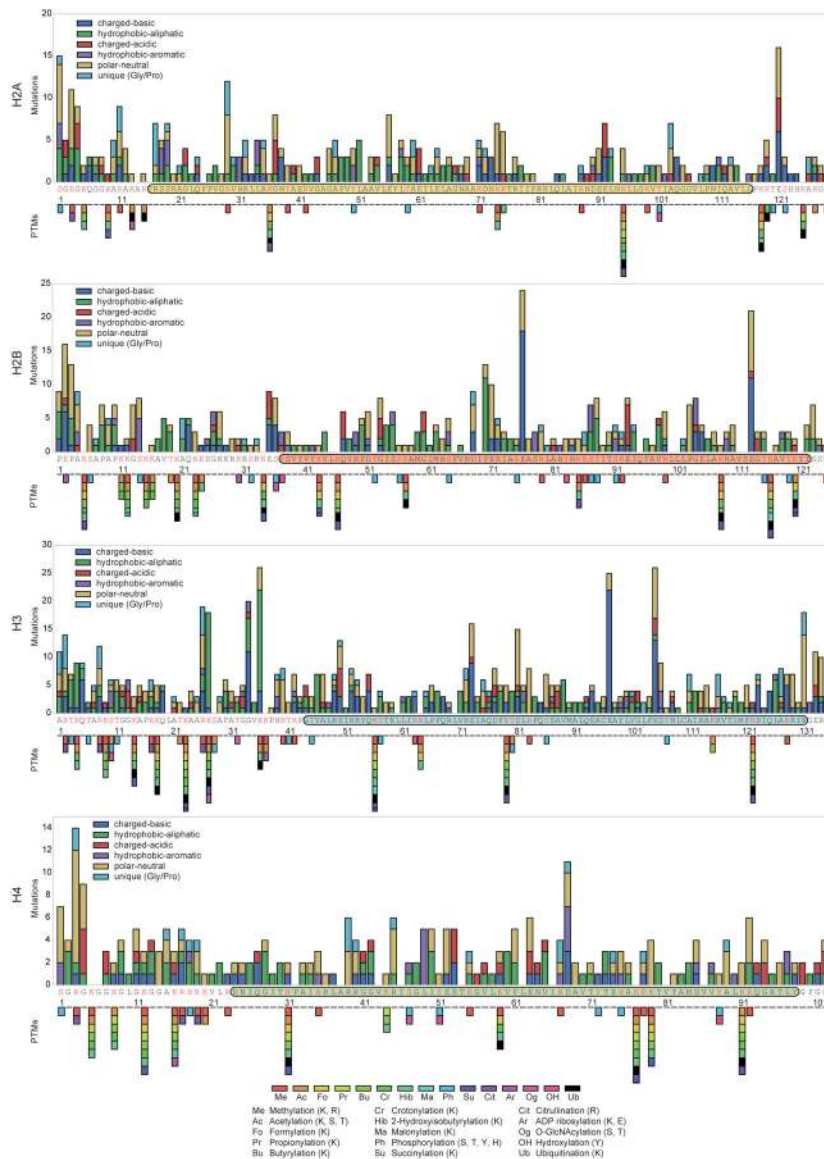
Extended Data Figure 4.
Heatmap of histone H4 mutations with individual residue labels. Color intensity indicates normalized mutation count (#mutations at residue/#samples per cancer type). Per patient data with all TMB is plotted. The numbers of tumors sequenced is indicated following the tumor type label.
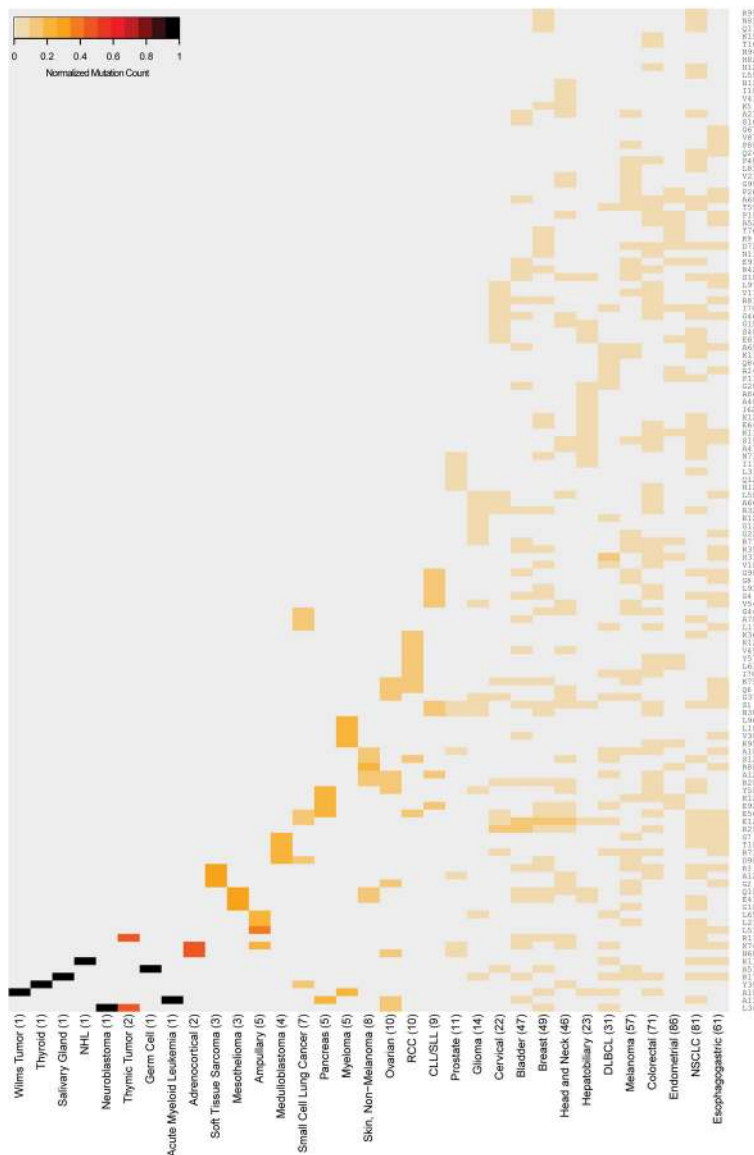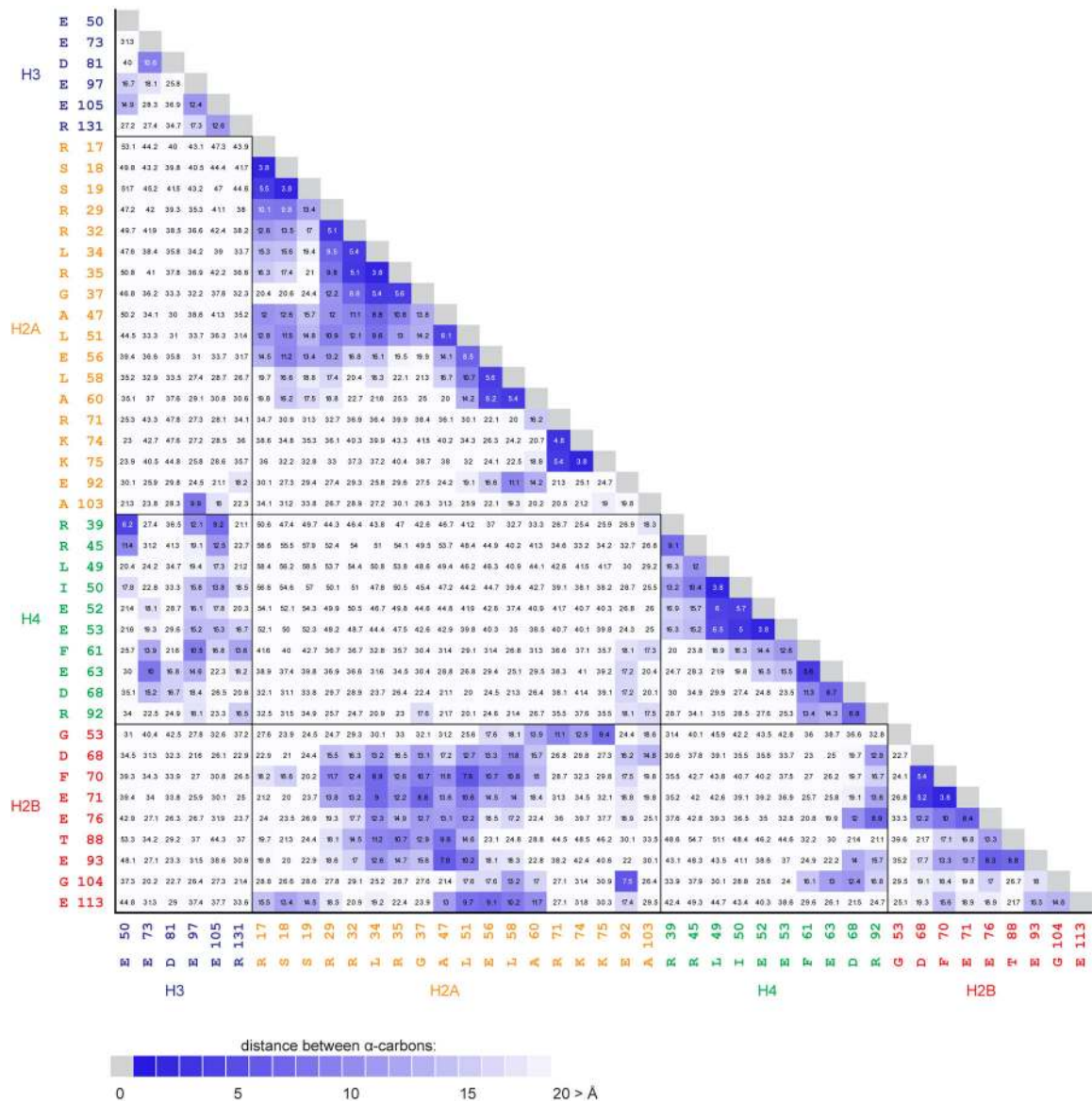
Extended Data Figure 5.
Heatmap of histone H2A mutations with individual residue labels. Color intensity indicates normalized mutation count (#mutations at residue/#samples per cancer type). Per patient data with all TMB is plotted. The numbers of tumors sequenced is indicated following the tumor type label.

Extended Data Figure 6.
Heatmap of histone H2B mutations with individual residue labels. Color intensity indicates normalized mutation count (#mutations at residue/#samples per cancer type). Per patient data with all TMB is plotted. The numbers of tumors sequenced is indicated following the tumor type label.
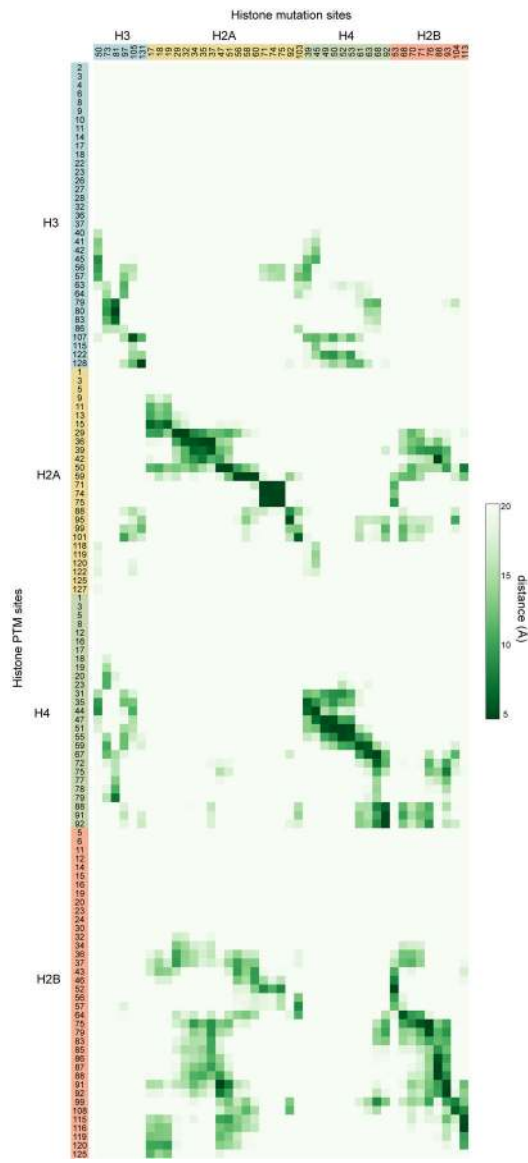
Extended Data Figure 7.
Histogram showing mutational frequency from the dataset in histones across all cancers. One tumor per patient where TMB ≤10 mut/Mb are shown. Boxes in the amino acid sequence show the globular domains of each histone. Amino acids with known post-translational modifications are marked in red, and the type of modification is shown by the bars below the histogram.
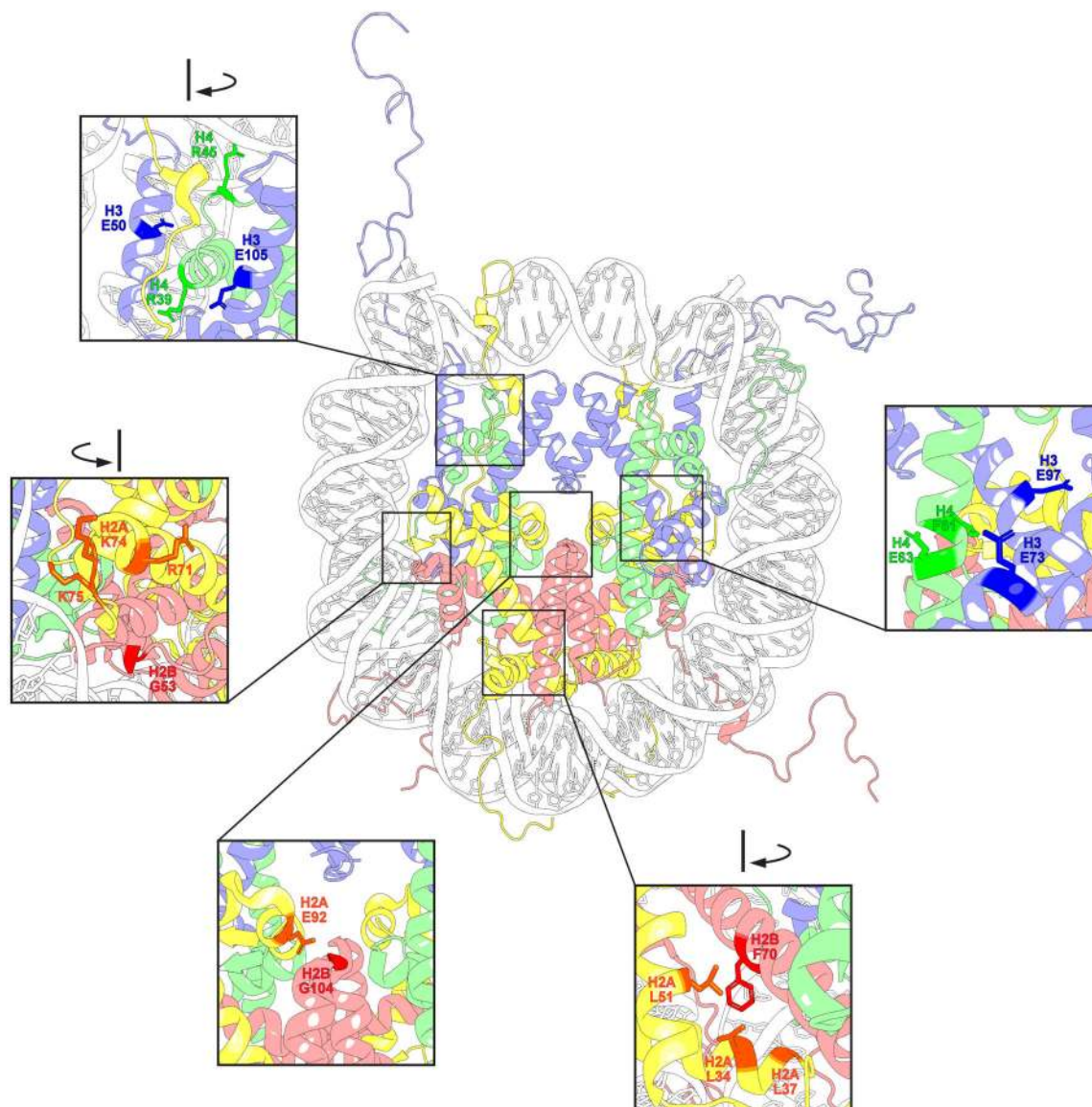
Extended Data Figure 8.
Proximity heat-map showing distances between the most frequently mutated residues in the nucleosome structure (PDB 1kx5). Samples with TMB ≤10 mut/Mb and mutation counts ≥ 2.5-fold the median number of mutations/residue for the histone family are displayed. Plotted residues are shown on the axes. Numbers within the grid indicate distance in angstroms between alpha-carbons.

Extended Data Figure 9.
Proximity heat-map showing distances between the most frequently mutated residues (horizontal axis) and sites of known PTMs (vertical axis). Per patient data at TMB ≤10 mut/Mb is shown for samples with mutation counts ≥2.5-fold the median mutations/residue for the histone family.

Extended Data Figure 10.

Frequently mutated residues converge in three-dimensional space. Examples of residues with alpha-carbons within 11.4Å that are mutated ≥2.5 fold over the median count/residue for each histone family when a TMB ≤10 mutations/Mb threshold is applied. Residues of interest are mapped on the nucleosome structure (PDB 1KX5).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References:

1. Luger K, Mäder a W., Richmond RK, Sargent DF & Richmond TJ Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260 (1997). [PubMed: 9305837]

2. Allis CD, Jenuwein T & Reinberg D Epigenetics Cold Spring Harb. Lab. Press. Cold Spring Harb New York (2007).

3. Strahl BD & Allis CD The language of covalent histone modifications. Nature 403, 41–45 (2000). [PubMed: 10638745]

4. Jenuwein T & Allis CD Allis: Translating the histone code. Science 293, 1074–1080 (2001). [PubMed: 11498575]

5. Shen H & Laird PW Interplay between the cancer genome and epigenome. Cell 153, 38–55 (2013). [PubMed: 23540689]

6. Schwartzentruber J et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature 482, 226–231 (2012). [PubMed: 22286061]

7. Behjati S et al. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat. Genet 45, 1479–82 (2013). [PubMed: 24162739]

8. Wu G et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet 44, 251–253 (2012). [PubMed: 22286216]

9. Lohr JG et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc. Natl. Acad. Sci. U. S. A 109, 3879–3884 (2012). [PubMed: 22343534]

10. Papillon-Cavanagh S et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat. Genet 49, 180–185 (2017). [PubMed: 28067913]

11. Zhao S et al. Mutational landscape of uterine and ovarian carcinosarcomas implicates histone genes in epithelial–mesenchymal transition. Proc. Natl. Acad. Sci 113, 12238–12243 (2016). [PubMed: 27791010]

12. Lewis PW et al. Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. Science 340, 857–861 (2013). [PubMed: 23539183]

13. Chan KM et al. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev 27, 985–990 (2013). [PubMed: 23603901]

14. Lu C et al. Cancer: Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 352, 844–849 (2016). [PubMed: 27174990]

15. Fang D et al. The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science 352, 1344–1348 (2016). [PubMed: 27229140]

16. Kandoth C et al. Mutational landscape and significance across 12 major cancer types. Nature 502, 333–339 (2013). [PubMed: 24132290]

17. Kruger W et al. Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. Genes Dev 9, 2770–2779 (1995). [PubMed: 7590252]

18. Gao J et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med 9, 4 (2017). [PubMed: 28115009]

19. Horn PJ, Crowley KA, Carruthers LM, Hansen JC & Peterson CL The SIN domain of the histone octamer is essential for intramolecular folding of nucleosomal arrays. Nat. Struct. Biol 9, 167–171 (2002). [PubMed: 11836537]

20. Lu P & Roberts CWM The SWI/SNF tumor suppressor complex: Regulation of promoter nucleosomes and beyond. Nucl. (United States) 4, (2013).

21. Hodges C, Kirkland JG & Crabtree GR The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. Cold Spring Harb. Perspect. Med 6, (2016).

22. McGinty RK & Tan S Recognition of the nucleosome by chromatin factors and enzymes. Current Opinion in Structural Biology 37, 54–61 (2016). [PubMed: 26764865]

23. Dann GP et al. ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. Nature 1–18 (2017). doi:10.1038/nature23671

24. Arimura Y 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 1–12 (2018). doi:10.1093/nar/gky661 [PubMed: 29177436]

25. Das C & Tyler JK Histone exchange and histone modifications during transcription and aging. Biochim. Biophys. Acta - Gene Regul. Mech 1819, 332–342 (2012).

26. Ye J et al. Histone H4 lysine 91 acetylation: A core domain modificationassociated with chromatin assembly. Mol. Cell 18, 123–130 (2005). [PubMed: 15808514]

27. Yan Q et al. BBAP Monoubiquitylates Histone H4 at Lysine 91 and Selectively Modulates the DNA Damage Response. Mol. Cell 36, 110–120 (2009). [PubMed: 19818714]

28. Tessadori F et al. Germline mutations affecting the histone H4 core cause a developmental syndrome by altering DNA damage response and cell cycle control. Nat. Genet 49, 1642–1646 (2017). [PubMed: 28920961]

29. Vogelstein B et al. Cancer genome landscapes. Science 339, 1546–58 (2013). [PubMed: 23539594]

30. Allis CD & Muir TW Spreading Chromatin into Chemical Biology. ChemBioChem 12, 264–279 (2011). [PubMed: 21243714]

31. Atlasi Y & Stunnenberg HG The interplay of epigenetic marks during stem cell differentiation and development. Nature Reviews Genetics 18, 643–658 (2017).

32. Cerami E et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov 2, 401–404 (2012). [PubMed: 22588877]

33. Gao J et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal Complementary Data Sources and Analysis Options. Sci Signal 6, 1–20 (2014).

34. Zehir A et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat. Med 23, 703–713 (2017). [PubMed: 28481359]

35. Cheng DT et al. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J. Mol. Diagnostics 17, 251–264 (2015).
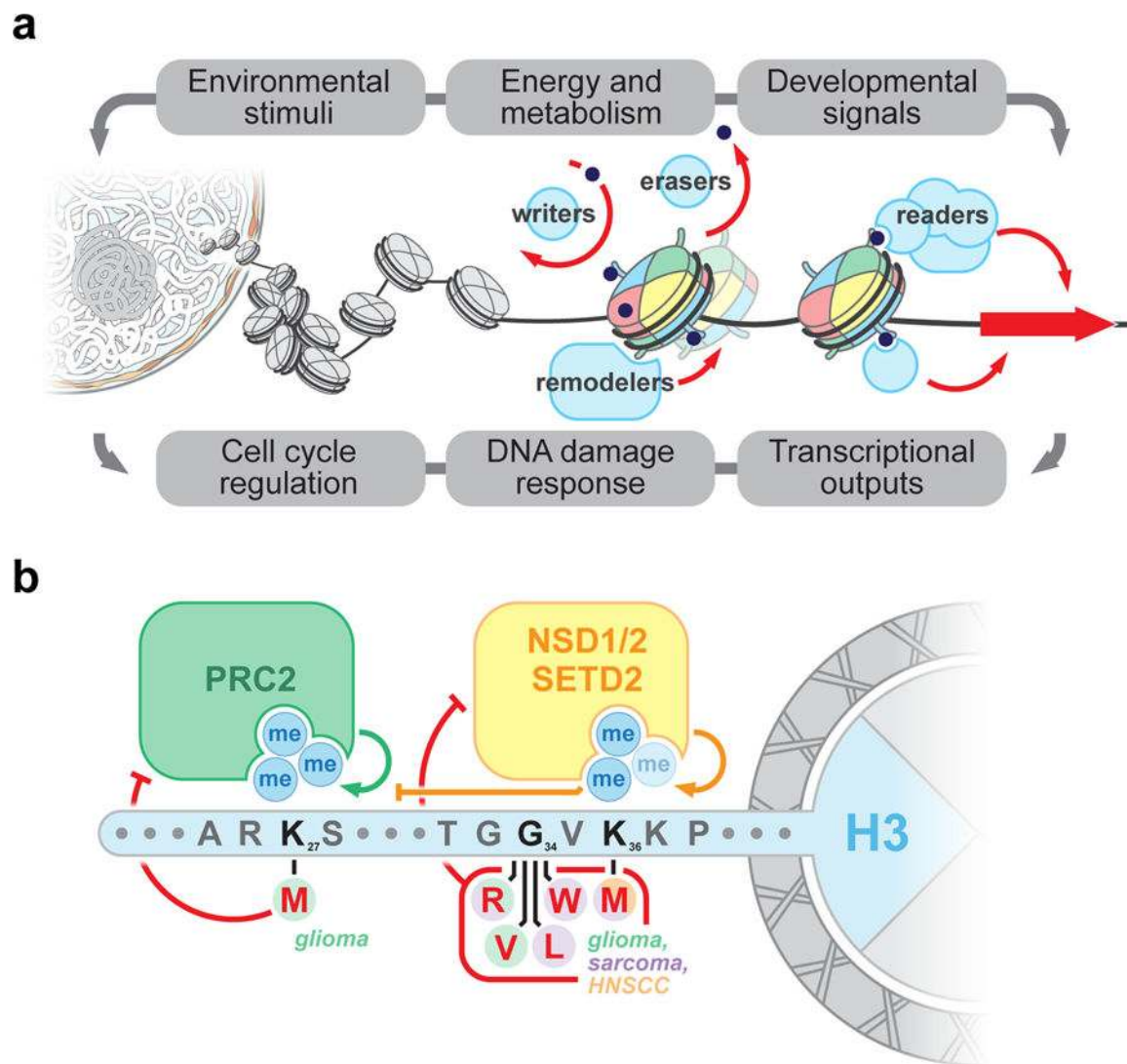
Figure 1. Histones as signal integrators and cancer driver genes.
a) Chromatin integrates environmental and developmental signals to control essential cell processes, including those dysregulated in cancer. b) Mechanisms and cancer type associations for known H3 oncohistone mutations.
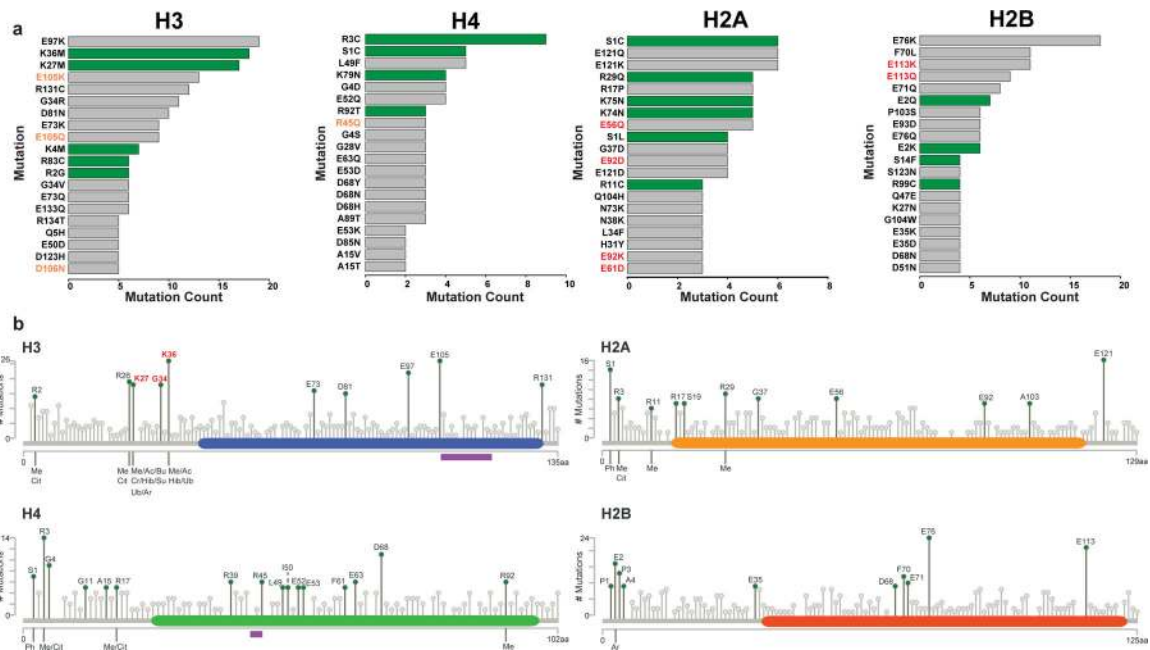
Figure 2. Cancer-associated histone mutations occur at sites of known PTMs and in both tail and globular domains.

a) The most prevalent somatic missense histone mutations for each core histone. Green bars, sites of known PTMs; orange lettering, residue in the 'Sin⁻' patch; red lettering, acidic patch residue. b) The 10 most frequently mutated residues in each core histone family shown in green/red; red labels, established oncohistone mutations. Globular domains are indicated by orange, blue, red, and green bars per color histone convention; purple bars, 'Sin⁻' patch. Type of PTMs are indicated below the domain structure schematic. See Extended Data Figure 7 for PTM legend.
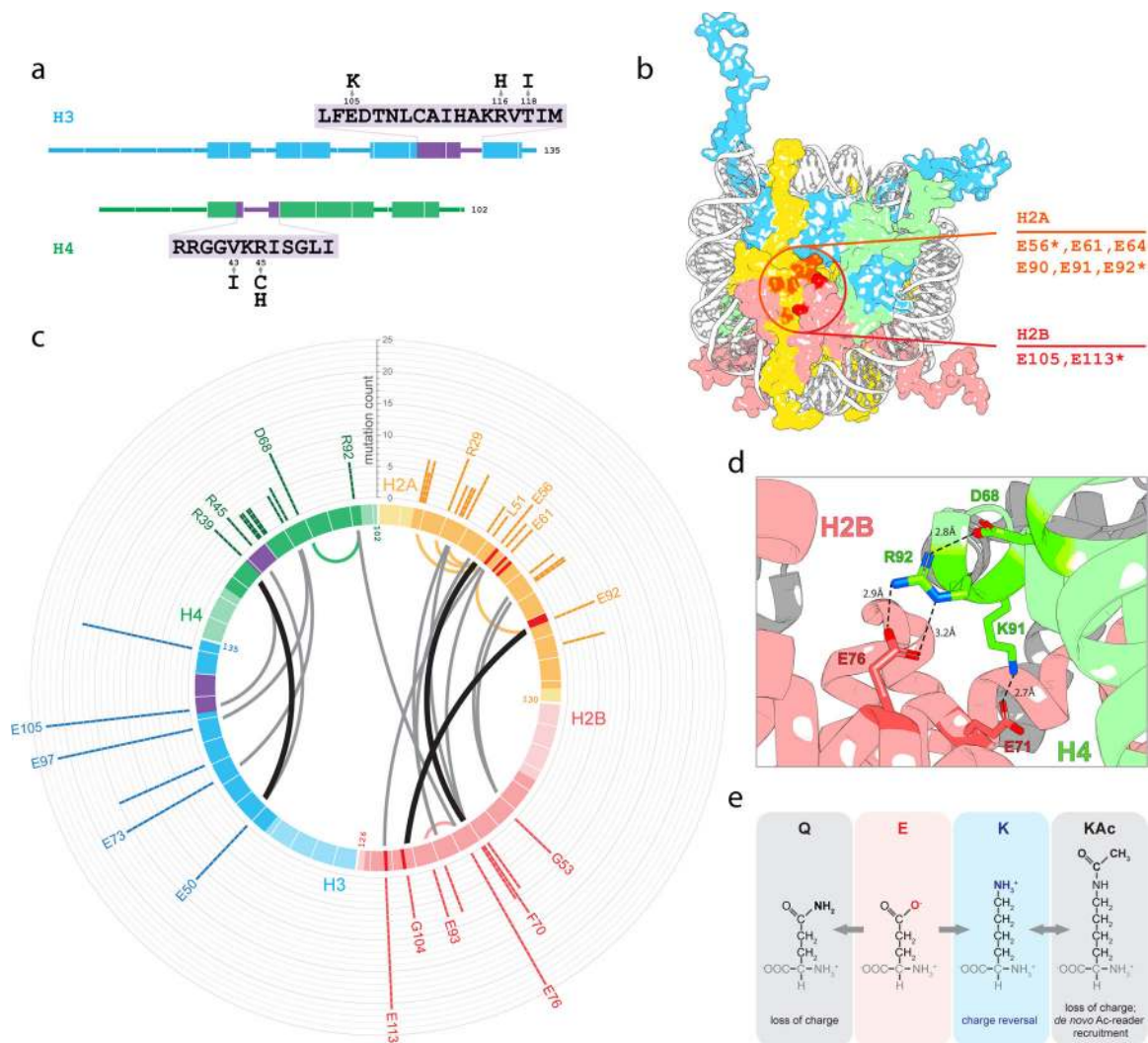
Figure 3. Hypothesis generating classes of histone mutations.
a) Histone mutations occur in the 'Sin-' patch. Boxes indicate alpha-helices. Classical yeast mutations are shown. b) Residues of the acidic patch (shown) are mutated in tumors. *Indicates amino acids with mutation counts ≥2.5-fold the median number of mutations for the core histone family. c) Mapping of three-dimensional proximity of residues mutated ≥ 2.5 fold over median mutation count/residue for each histone family on the nucleosome structure (PDB 1KX5). Globular domains are shaded in darker hues, Sin- patches in purple, and acidic patches in red. Bar height indicates mutation count. Thick black lines indicate alpha-carbon distance between 3.8 and 7.6Å and think grey lines indicate 7.6 to 11.4Å. Intra-molecular proximities are indicated by colored lines and only select residues are labeled for clarity of display. d) The H2B-H4 interface is mediated by hydrogen bonding between H2BE76, H4D68 and H4R92, as well as a salt bridge between H2BE71 and H4K91. These residues exhibit high mutational frequency except H4K91. e) Glutamic acid residues are frequently mutated to lysines or glutamines, which can serve as substrates for acetylation or function as acetyl mimics, respectively.
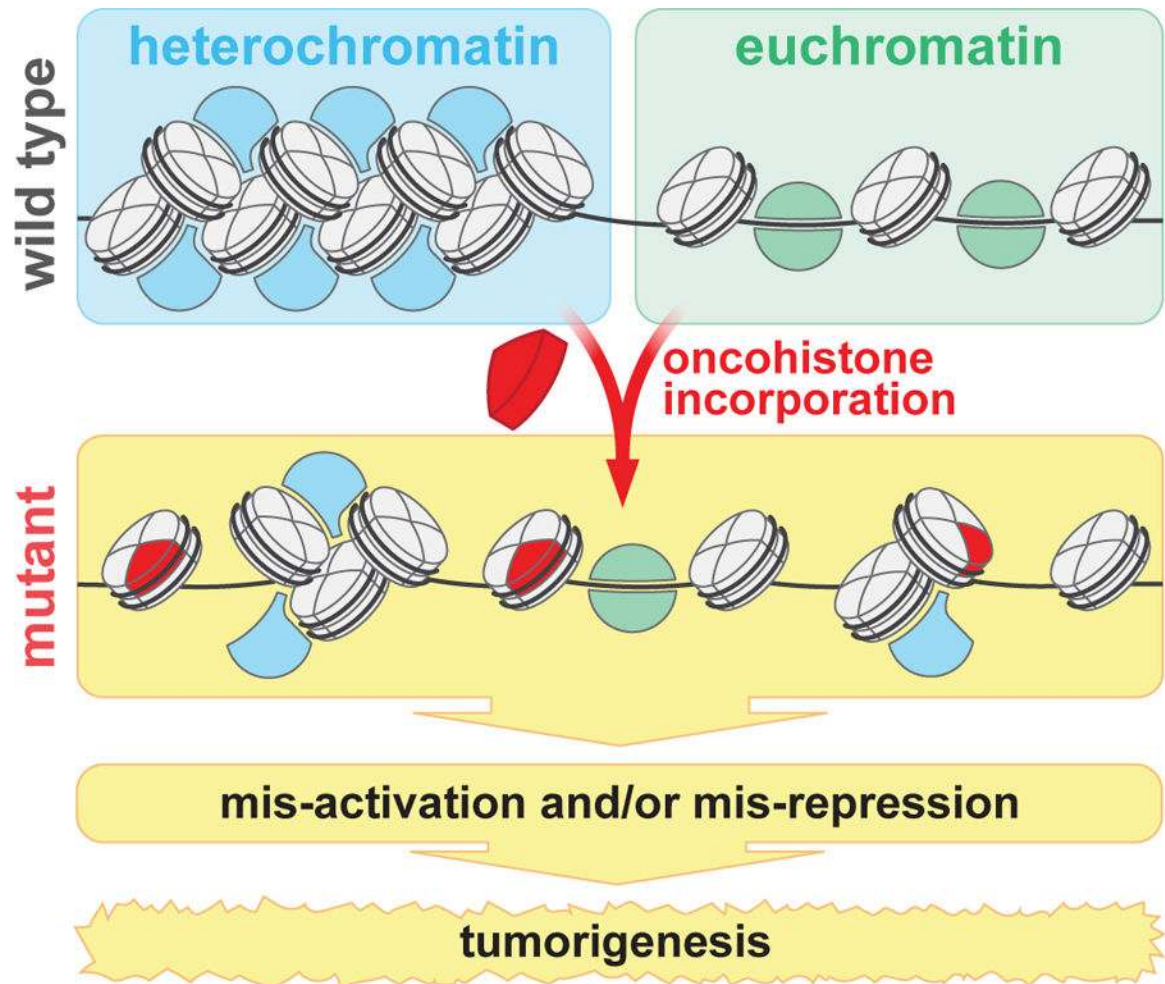
Figure 4. A model for the impact of oncohistones on the chromatin polymer.
By incorporation into the chromatin polymer, mutated histone proteins (oncohistones, in red) may cause functional effects by altering the biophysical and/or functional properties of chromatin. We propose these effects will occur when the mutant histone is present even at low concentrations and that mutating even one of many histone gene copies can have dominant effects.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

The most prevalent mutated histone residues in the dataset

Mutated Residue	Mutation Count	Histone	Domain
E105	26	H3	Globular
K36	26	H3	N-terminal
E76	24	H2B	Globular
E97	22	H3	Globular
E113	21	H2B	Globular
G34	20	H3	N-terminal
R26	19	H3	N-terminal
K27	18	H3	N-terminal
R131	18	H3	Globular
E73	16	H3	Globular
E121	16	H2A	C-terminal

Table 2.

Residues with potentially functional mutations by 3D hotspot analysis[18]

Histone Family	Uniprot ID	Residues	Structure	P-value
H2A	H2A1A_HUMAN	27, 29, 32, 34, 35	5kgf.G	0.081
H2A	H2A1H_HUMAN	29, 30, 33, 35, 37, 38	1p3b.C	0.029
H2B	H2B1L_HUMAN	73, 89, 90, 93, 94, 98, 99	3b6g.D	0.02
H3.1	H31_HUMAN	2 – 8	swissmodel:5wvo.D	0.005
H3.1	H31_HUMAN	101 – 109,127,131,132,134,135	5kdm.A	0.098
H4	H4_HUMAN	2 – 4, 7 – 15	1kx5.B	0.094