A common classification framework for histone sequence alterations in tumours: an expert consensus proposal

Henning Leske1,2*, Raymond Dalgleish3, Alexander J Lazar4, Guido Reifenberger5 and Ian A Cree6

1 Department of Pathology, Oslo University Hospital, Oslo, Norway
2 University of Oslo (UiO), Oslo, Norway
3 Department of Genetics and Genome Biology, University of Leicester, Leicester, UK
4 Departments of Pathology & Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
5 Institute of Neuropathology, Heinrich Heine University, Medical Faculty, Düsseldorf, Germany
6 International Agency for Research on Cancer (IARC), World Health Organization, Lyon, France

*Correspondence to: H Leske, Department of Pathology, University Hospital Oslo, Sognsvannsveien 20, 0372 Oslo, Norway.
E-mail: henning.leske@medisin.uio.no; henning.leske82@gmail.com

Abstract

The description of genetic alterations in tumours is of increasing importance. In human genetics, and in pathology reports, sequence alterations are given using the human genome variation society (HGVS) guidelines for the description of such variants. However, there is less adherence to these guidelines for sequence variations in histone genes. Due to early cleavage of the N-terminal methionine in most histones, the description of histone sequence alterations follows their own nomenclature and differs from the HGVS-compliant numbering by omitting this first amino acid. Next generation sequencing reports, however, follow the HGVS guidelines and as a result, an unambiguous description of sequence variants in histones cannot be provided. The coexistence of these two nomenclatures leads to confusions for pathologists, oncologists, and researchers. This review provides an overview of tumour entities with sequence alterations of the H3–3A gene (HGNC ID = HGNC:4764), highlights the problems associated with the coexistence of these two nomenclatures, and proposes a standard for the reporting of histone sequence variants that allows an unambiguous description of these variants according to HGVS principles. We hope that scientific journals will adopt the new notation, and that both geneticists and pathologists will include it in their reports.

Keywords: histone; classification; mutation; tumour; neoplasm; nomenclature; sequence

Received 28 January 2021; Revised 17 March 2021; Accepted 25 March 2021

No conflicts of interest were declared.

Introduction

The diagnosis and management of tumours increasingly depends on the demonstration of sequence variants (colloquially known as mutations and other genetic alterations) in genes related to their causation. The discovery of such tumour-associated sequence variants not only facilitates the diagnosis of tumour types that are difficult to classify based on histomorphology and immunohistochemistry alone but can also be of prognostic and therapeutic value. To unify the description of such variants and to make new discoveries available for clinicians and pathologists, the Human Genome Variation Society (HGVS) has developed and maintained a nomenclature for the standardised and unambiguous description of sequence variants [1–4]. This has become the universally accepted nomenclature for all next generation sequencing (NGS) and whole genome sequencing reports and provides the basis for consensus recommendations for molecular pathology [5] and clinical oncology [6].

Biologists and clinicians around the world have applied these guidelines to report new discoveries in DNA sequence variation, with the apparent exception of those working on histones, which present unique challenges.

Histones are a family of basic proteins that associate with DNA and allow condensation to form chromatin. The four canonical histones H2A, H2B, H3, and H4 form dimers between H2A and H2B (H2A–H2B) and between H3 and H4 (H3–H4). When pairs of these dimers are bound to DNA, they form a histone octamer, which corresponds to a nucleosome. A chain of nucleosomes is wrapped in a spiral called a solenoid, in which the nucleosomes are further stabilised by the linker histone H1 to support the chromatin structure [7,8].

Histone post-translational modifications such as acetylation, phosphorylation, methylation, ubiquitylation, and sumoylation, as well as histone exchange, affect
DNA transcription by regulating RNA polymerase II’s access to DNA. Consequently, complex cell processes can be influenced by histone modifications and its related alterations in gene expression. There are isoforms for the core histones H3, H2A, and H2B, as well as for the H1 linker histone, which show alterations of either a few amino acids or larger domains [7]. These histone isoforms are further subdivided into replicative or replacement types. Whereas replicative types are only highly expressed during the S-phase, replacement histone types are typically synthesised in a cell-cycle-independent manner and incorporated into chromatin. Since replacement histone types localise to specific chromatin domains, they can lead to differential expression in different tissue compartments. Therefore, a single histone isoform can fulfil different functions. Replacement types H3.3 and H2AZ are involved in the regulation of transcription, whereas the replicative isoforms H3.1 and H3.2 are incorporated during replication and the replicative H3.5 contributes to DNA condensation during spermatogenesis [9,10].

Well-documented discrepancies in histone sequence notation reflect the historical roots of this scientific field. The amino acid (AA) sequence of histones was first deciphered in the 1960s and 1970s [11–13], using chromatographic separation methods, enzymatic and chemical cleavage, followed by Edman degradation to determine the sequences of the functional histone proteins. After histone genes had been sequenced, it became apparent that the AA encoded by the first codon of histone genes, a methionine, is post-translationally cleaved from most proteins and could therefore not be detected with previous protein-based methods. However, the previously established numbering of the AA sequences remained the standard in this field. Therefore, the sites of histone sequence variants are typically reported lacking one AA, relative to the DNA-based AA numbering [11,14]. This discrepancy is particularly problematic at sites where two identical AAs follow each other in the protein sequence. In human H3.3 histone A (encoded by the H3-3A gene; HGNC ID = HGNC:4764), such DNA-based numbering sites are at positions 13/14, 25/26, 34/35, 37/38, 53/54, 61/62, 88/89, and 129/130. Unfortunately, one of the crucial sites widely accepted as an oncogenic driver is the H3.3 histone A sequence variant at AA 34/35, depending on the numbering scheme being used. Consequently, it is unclear which AA is meant, because in the HGVS-compliant protein reference sequence (NCBI RefSeq NP_002107.7) there is a glycine at both AA 34 and AA 35. Thus, the histone legacy AA variant H3.3 G34 corresponds to a DNA-based H3.3 G35 variant using current variant reporting standards, as would be reported by current NGS data pipelines [15].

While in some instances the use of legacy AA numbering may have little significance, it is not sufficient for use in tumour classification. For histone sequence variants, the coexistence of two numbering systems (nomenclatures) can lead to confusion for pathologists, researchers, and clinicians alike. The International Agency for Research on Cancer (IARC) noted this issue during production of the 5th edition of the WHO Classification of Tumours, published as the widely used WHO Blue Books (http:// whobluebooks.iarc.fr), and decided to explore the matter further in the hope of providing precise and unequivocal description of tumour-associated sequence variants in histone genes to avoid the potential for errors based on the coexistence of two different nomenclatures.

We therefore propose the use of the following notation: in line with existing guidance for reporting genetic alterations, histone sequence variants should be reported using the HGVS recommended format. First, reference sequences must be specified for the transcript (mRNA) and the corresponding encoded protein. Once that has been done, the gene symbol may be used to indicate the gene that harbours the DNA sequence. For example, the transcript and protein reference sequences for the H3-3A gene are RefSeq [16] entries NM_002107.7 and NP_002098.1, respectively. Based on these reference sequences, a variant in H3-3A at the transcript and protein level would be described as H3-3A:c.103G>A p.Gly35Arg. This variant description is standards-compliant and corresponds to the variant commonly known as H3.3 G34R, in which the AA number is based on the legacy protein sequence and the amino acids are demonstrated in a single letter code. This legacy designation is also used to label antibodies raised against specific protein isoforms and variants. It should be noted that there is currently no agreed HGVS nomenclature syntax for the reporting of legacy/common variant descriptions. For now, the reporting format from DNA sequencing should be H3-3A:c.103G>A p.Gly35Arg (G34R), noting that there is a space, but no punctuation, between the DNA, protein, and the legacy protein variant descriptions, often identified by immunohistochemistry. The parentheses are used to indicate the legacy protein sequence description.

The WHO Classification of Tumours will adopt this nomenclature within its publications both online and in print. Molecular pathology reports should also cite the version of the human genome used, e.g. GRCh37 or GRCh38, when a variant is described in terms of a genome position following genome or exome sequencing, as well as the method used to obtain the results [17,18].

We established the extent of the nomenclature problem for sequence variants of histones using these conditions, parameters, and assumptions:

- Nucleotide and AA sequence variants are described in accordance with the current HGVS recommendations (Version 20.05) for the description of sequence variants [1]. In line with these recommendations, the term ‘sequence variant’ is used instead of ‘mutation’. In agreement with current recommendations, the three-letter AA code is used to reduce reporting errors. To simplify the presentation of HGVS-compliant protein sequence variant descriptions, an assumption is made that all such variants have been confirmed by independent methods and that none are purely predictions from variants at the nucleotide level.

© 2021 The Authors. The Journal of Pathology published by John Wiley & Sons, Ltd. on behalf of The Pathological Society of Great Britain and Ireland. www.pathsoc.org
Tumour entities that harbour histone-related changes

The discovery of histone sequence variants as oncogenic drivers in neoplasms has opened a new chapter in tumour diagnostics. Indeed, there are an increasing number of tumour entities that harbour histone-related changes (Table 2). Such entities include tumours of the central nervous system, bone and soft tissue neoplasms, head and neck squamous cell carcinoma, malignant melanoma, bladder and colorectal cancer, ovarian cancer, and haematological neoplasms. These are described here and summarised in Table 3.

Digestive system tumours

Few H3-3A sequence variants have been reported in COSMIC (H3F3A) for digestive system tumours, and there is as yet little consistency, with fewer than five reports of the same tumour type. Though these are small numbers of reports, the few anatomic tumour sites that carry alterations in more than one case comprise H3-3A sequence variants encoding p.Ala115Gly, p.Ala88Ser, p.Ala88Thr, p.Ile90Val, p.Lys28Arg, p.Arg50His, or p.Arg50Cys missense variants. Such changes have been detected in intestinal adenocarcinoma, tumours involving the oesophagus and pancreas, as well as hepatocellular cholangiocarcinoma [34]. The biological significance of these sequence alterations is uncertain.

Breast tumours

There is a high diversity of H3-3A sequence variants described for breast tumours in the COSMIC database, most being invasive breast carcinomas (IBCs) of no special type (NST). However, COSMIC lists two ductal carcinomas in situ in combination with Paget’s disease, which both show a p.Arg73Gln substitution.

Soft tissue and bone tumours

There is an established and consistent pattern of sequence alteration in giant cell tumour of bone [23,26,35], with the H3-3A:c.103G>T p.Gly35Trp sequence variant being reported in the vast majority of cases. Additionally, chondroblastoma, chondrosarcoma, clear cell chondrosarcoma, and osteosarcoma have been reported to harbour H3-3A: c.110A>T p.Lys37Met sequence variants [23,36].

Female genital tumours

There are few reports of H3-3A gene sequence alterations in tumours of the female genital tract, with most reported in endometrial carcinomas (18/24 in COSMIC; supplementary material, Table S1) [37].

Thoracic tumours

There are few reports of sequence variants in thoracic tumours. Two small cell lung carcinomas have been reported to harbour sequence variants leading to a stop at p.Gly34Ter [38]. Squamous cell carcinoma cases of the lung have been described with the introduction of stop codons at p.Gln6Ter or p.Tyr100Ter (CGP study 418).

Central nervous system tumours

The majority of H3-3A sequence variants have been reported in tumours of the CNS, some of which are now named accordingly:

Diffuse hemispheric glioma, H3 G34–mutant

The incidence of diffuse hemispheric glioma, H3 G34-mutant, ranges from 8% to 16% of paediatric high-grade gliomas with hemispheric location [25,29,39]. So far, H3-3A sequence alterations resulting in a p.Gly35Arg (G34R) or a p.Gly35Val (G34V) substitution have been described for this tumour type. The vast majority of tumours show a glycine-to-arginine substitution (p.Gly35Arg; 94%),

© 2021 The Authors. The Journal of Pathology published by John Wiley & Sons, Ltd. on behalf of The Pathological Society of Great Britain and Ireland. www.pathsoc.org
whereas the glycine-to-valine substitution is rare (p.Gly35Val; 6%).

Diffuse midline glioma, H3 K27–altered
In children, diffuse midline gliomas, H3 K27–altered, are preferentially located in the brainstem, particularly in the pons, also known as diffuse intrinsic pontine glioma (DIPG), whereas diffuse midline gliomas, H3 K27–altered, in adults are more commonly seen in the thalamus or spinal cord [40,41]. The vast majority of H3 K27–altered diffuse midline gliomas contain the p.Lys28Met (K27M) missense change in one of the histone H3 isoforms [25,42], while individual cases of DIPG may carry a p.Lys28Ile (K27I) substitution [24]. There have been several histone genes reported that may contain analogous sequence variants, with H3-3A being most commonly affected, while H3-3B, H3C2, H3C3, and H3C14 variants are less common and mostly restricted to DIPG [24,25,42].

Pilocytic astrocytoma (with histone H3 gene alteration)
Pilocytic astrocytoma carrying a histone H3 sequence alteration that results in a p.Lys28Met (K27M) substitution is extremely rare. Some of the reported cases showed fast tumour progression and hence are more likely to resemble diffuse midline gliomas, H3 K27–altered. However, there are individual cases with a prolonged overall survival of about 10 years [43], which are biologically more in accordance with pilocytic astrocytomas.
Table 2. Tumour-associated sequence variants of histones.

| HGVS gene symbol (HGNC ID) | Previous gene symbol | NCBI RefSeq reference transcript sequence | Transcript-level variant description | NCBI RefSeq reference protein sequence | Protein-level variant description | Legacy protein-level variant description* | PMID |
|---------------------------|----------------------|------------------------------------------|-------------------------------------|---------------------------------------|-----------------------------------|------------------------------------------|-------|
| H1-6 [HGNC:4720]          | HIST1H1T             | NM_005323.4 c.426G>C                    | c.426G>C                            | NP_005314.2                            | p.Lys142Asn                        | K141N                                    | [22]  |
| H3C10 [HGNC:4775]         | HIST1H3H             | NM_003536.3 c.37G>C                      | c.37G>C                             | NP_003527.1                            | p.Gly13Arg                         | G12R                                     | [23]  |
| H3C2 [HGNC:4776]          | HIST1H3B             | NM_003537.4 c.83A>T                      | c.83A>T                             | NP_003528.1                            | p.Lys28Met                         | K27M                                     | [24]  |
| H3C3 [HGNC:4768]          | HIST1H3C             | NM_003531.3 c.83A>T                      | c.83A>T                             | NP_003522.1                            | p.Lys28Met                         | K27M                                     | [24]  |
| H3C14 [HGNC:20503]        | HIST2H3C             | NM_020109.3 c.83A>T                      | c.83A>T                             | NP_066403.2                            | p.Lys28Met                         | K27M                                     | [24]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.83A>T                      | c.83A>T                             | NP_002098.1                            | p.Lys28Met                         | K27M                                     | [25]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.83_84delinsTT              | c.83_84delinsTT                      | NP_002098.1                            | p.Lys28III                         | K27I                                     | [24]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.104G>A                     | c.104G>A                            | NP_002098.1                            | p.Gly35Glu                         | G34E                                     | [26]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.103_104delinsCT            | c.103_104delinsCT                    | NP_002098.1                            | p.Gly35Leu                         | G34L                                     | [26]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.103_104delinsAT            | c.103_104delinsAT                    | NP_002098.1                            | p.Gly35Met                         | G34M                                     | [26]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.103G>A                     | c.103G>A                            | NP_002098.1                            | p.Gly35Arg                         | G34R                                     | [25]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.103G>C                     | c.103G>C                            | NP_002098.1                            | p.Gly35Arg                         | G34R                                     | [27]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.104G>T                     | c.104G>T                            | NP_002098.1                            | p.Gly35Val                         | G34V                                     | [25]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.103G>T                     | c.103G>T                            | NP_002098.1                            | p.Gly35Trp                         | G34W                                     | [23]  |
| H3-3A [HGNC:4765]         | H3F3B                | NM_005324.5 c.103G>A                     | c.103G>A                            | NP_005315.1                            | p.Gly35Arg                         | G34R                                     | [23]  |
| H3-3A [HGNC:4764]         | H3F3A                | NM_002107.7 c.110A>T                     | c.110A>T                            | NP_002098.1                            | p.Lys37Met                         | K36M                                     | [23]  |
| H3-3B [HGNC:4765]         | H3F3B                | NM_005324.5 c.110A>T                     | c.110A>T                            | NP_005315.1                            | p.Lys37Met                         | K36M                                     | [23]  |
| H4-16 [HGNC:20510]        | HIST4H4              | NM_175054.2 c.167G>A                     | c.167G>A                            | NP_778224.1                            | p.Arg66Gln                         | R55Q                                     | [22]  |

*The legacy protein-level variant descriptions are based upon amino acid numbering of the mature histone protein following post-translational removal of the initiating methionine, as used in COSMIC v91.

Ganglioglioma (with histone H3 gene alteration)

Histone sequence variants in gangliogliomas are rare. Reported cases with histone H3 p.Lys28Met (K27M) alteration are, so far, associated with a midline location of the tumour. Such sequence variants have been more frequently described in anaplastic gangliogliomas (WHO CNS grade 3) rather than in classic gangliogliomas (WHO CNS grade 1) [44]. The outcome for these patients has been reported to be poor, even in the absence of signs of anaplasia, although there are patients with prolonged overall survival [44,45]. There are single reports showing that diffuse hemispheric gliomas with the H3-3A:c.103G>A p.Gly35Arg (G34R) variant may contain a dysplastic neuronal component, as typically seen in gangliogliomas. Whether these resemble a morphological variant of diffuse hemispheric glioma, H3 G34-mutant, or whether they are true gangliogliomas is, to date, unclear and requires further evaluation [46].

Posterior fossa ependymoma group PFA (with histone H3 gene alteration)

Posterior fossa ependymomas group A (PFA) exhibit EZH inhibitory protein (EZHIP) overexpression that may mimic the biological effects of the histone H3 p.Lys28Met (K27M) variant [47]. Only exceptional cases of posterior fossa ependymomas group A show an H3-3A:c.83A>T p.Lys28Met (K27M) sequence alteration [48,49]. The progression-free survival of PFA cases is worse compared with its counterpart, the posterior fossa ependymoma group B, which lacks sequence alterations of one of the histone H3 isoforms [50].

Subependymoma (with histone H3 gene alteration)

A single study reported on individual brainstem subependymomas with H3-3A:c.83A>T p.Lys28Met (K27M) alteration and long overall survival [51].
Table 3. Individual tumour types with histone H3.3 sequence variants described in COSMIC, and selected references. Tumours with fewer than five entries containing the same H3-3A sequence variants in COSMIC (supplementary material, Table S1) are excluded.

| Tumour type | Histone sequence variants | Reference |
|-------------|---------------------------|-----------|
| 1. Digestive system tumours | Fewer than five entries |          |
| 2. Breast tumours | Fewer than five entries |          |
| 3. Soft tissue and bone tumours | Fewer than five entries |          |
| 4. Female genital tumours | Fewer than five entries |          |
| 5. Tumours of the lung, pleura, thymus and heart | Fewer than five entries |          |
| 6. Tumours of the central nervous system | Fewer than five entries |          |
| 7. Tumours of the urinary system and male genital organs | Fewer than five entries |          |
| 8. Head and neck tumours | Fewer than five entries |          |
| 9. Tumours of endocrine organs | Fewer than five entries |          |
| 10. Tumours of haematopoietic and lymphoid tissues | Fewer than five entries |          |
| 11. Skin tumours | Fewer than five entries |          |
| 12. Tumours of the eye | Fewer than five entries |          |

Meningiomas

Two intronic H3-3A variants have been described in rare cases of intraventricular meningiomas, H3-3A:c.103G>T and H3-3A:c.104G>T [28]. The biological impact of these alterations remains elusive.

Urinary and male genital tumours

There are relatively few reports of H3-3A sequence variants in prostate, bladder, and renal carcinomas [34], with H3.3 histone A p.Ala115Gly, p.Lys37Met, p.Arg50His and p.Arg50Leu protein alterations each being listed at least twice (supplementary material, Table S1).

Head and neck tumours

Squamous cell carcinomas of the upper respiratory tract [34,53] have been reported to carry a recurrent H3-3A: c.344C>G p.Ala115Gly alteration, though numbers remain small. So far, these alterations have only been found in HPV-independent tumours.

Endocrine and neuroendocrine tumours

Histone sequence variants have been studied in relatively small numbers of cases, but there is a consistent H3-3A: c.103G>T p.Gly35Trp variant in four phaeochromocytomas listed in COSMIC, and in four paragangliomas studied [22]. This variant also occurs in thyroid carcinomas (2/4 cases, type not specified), whereas the other two cases of H3-3A-mutant thyroid cancer displayed an H3-3A: c.86G>C p.Ser29Thr variant (CGP study 676).

Haematolymphoid tumours

Histone sequence variants occur in a variety of myeloid and lymphocytic haematogenous malignancies, with T-cell lymphomas showing a p.Ala115Gly, p.Lys28Glu, p.Lys28Met or p.Lys28Asn substitution in H3.3 histone A in more than one case. A single case of a diffuse large B-cell-lymphoma with an H3-3A:c.45A>C p.Lys15Asn substitution has been described (CGP 632).

Lohr et al [54] reported an accumulation of variants in the linker histone H1 family in diffuse large B-cell lymphomas. The functional significance of these sequence alterations remains to be explored, although the authors report that hotspot variants of H3C2 and potentially other histone core proteins might be related to activation-induced cytidine deaminase (AID)-mediated somatic hypermutation [54].

Additionally, single leukaemia cases including AML [34] and T-ALL [55,56], as well as follicular lymphomas [57], have been reported to show alterations in histone genes.

Skin tumours

There are rare cases of melanomas with H3-3A sequence variants [34].

Advantages of using the proposed reporting system for histone sequence alterations

Our analysis of the published literature and the COSMIC database suggests that H3-3A sequence variants occur in tumours from various sites of the body, though overall relatively rarely, with CNS and bone tumours accounting for the majority of reported cases.

The use of a standardised nomenclature for reporting genetic alterations, as described in the HGVS Sequence Variant Nomenclature, allows the unambiguous description of sequence variants, independent of the analytical method, tumour type or field of research. The HGVS recommendations have thus become the almost universally accepted guidelines for the description of genetic alterations.

The benefit of the recommended HGVS-compliant reporting scheme is that it includes the sequence alteration at the DNA level, as well as the predicted resulting change at the protein level. This is particularly relevant for histones for the following reasons. Histone H3.3 is encoded by two genes, H3-3A and H3-3B, with each
gene encoding identical protein products despite the
genes themselves having distinct and different nucleo-
tide sequences. Sequence variants can occur in either of
these genes, resulting in the same altered protein prod-
uct. For example, in chondroblastomas [58], the poten-
tial p.Lys37Met substitution of histone H3.3 results
from sequence variants in both the H3-3A and the
H3-3B genes, with the majority of these cases showing
the sequence variation in H3-3B [23] (supplementary
material, Figure S1). An ‘H3.3 K37M’ description
would thus not be sufficient to unequivocally report the
differences at the site of genomic sequence alteration.

The report of the specific mutated gene or alteration
may also have prognostic impact, as observed in paediat-
ric patients with H3 K27-altered diffuse midline glioma
located in the pons, i.e. diffuse intrinsic pontine glioma
(DIPG). DIPG patients with an H3-3A:c.83A>T p.
Lys28Met (K27M) sequence variant have a median
overall survival of 11 months. In contrast, patients with
the same AA substitution resulting from sequence varia-
tion in the H3C2, H3C3 or H3C14 genes have a median
overall survival of 15 months [59–61]. Since this infor-
mation can be useful for clinicians and patients, the
recommended HGVS-compliant nomenclature encour-
gages pathologists to correctly report alterations that
are associated with different outcomes. Such information
is important for future studies to unequivocally report
 genetic information that is known to be of prognostic rel-
evance. Additionally, it enables better-defined patient
stratification that might result in the identification of
new prognostically or therapeutically relevant sequence
alterations in the future.

Despite the clearly defined site of sequence variation
and report of potentially different outcome in patients,
the so far-used AA-based description of the site of his-
tone sequence alteration is potentially inaccurate. The
legacy description (H3 K27) in ‘diffuse midline gli-
oma, H3 K27-altered’ is based on the processed his-
tone protein, after cleavage of the initiating methionine, which is assumed to be the tumour driver
protein. However, histones contain several N-terminal
cleavage sites. For histone H3 variants, several cathepsin
L cleavage sites have been reported [62] in murine
and human embryonic stem cells (ESCs), such as after
p.Ala22, p.Thr23, p.Lys24, p.Ala25, p.Arg27 and
p.Lys28. Mass spectrometry analysis has demonstrated
that such post-translational cleavage occurs. Even
though the role of such cleavage has not been entirely
resolved, Zhou et al [62] reported that it may influence
the epigenetic signature upon differentiation and that it
could play a role in apoptosis as well as in immune
cell recruitment. Alterations of these hypothesised
functions are tightly associated with tumour develop-
ment [63]. In addition, there has been a serine protease
activity detected in human ESCs with an additional
cleavage site after p.Ala32 in H3.1 and/or H3.2 iso-
forms [64]. To our knowledge, this serine protease
has not been fully characterised and thus we lack fur-
ther information regarding the biological impact. The
use of an AA-based nomenclature, as is the current
standard in the histone field, is not only potentially
ill-defined but also potentially erroneous regarding
the tumour driver protein, with implications for both
diagnostics and research.

Another aspect is that there is at least one histone (H2B
type W-T) in which the initiating methionine is not
cleaved. Therefore, the histone nomenclature description
of sequence variants necessitates awareness of whether
cleavage of the initiating methionine occurs [7].

It becomes even more complicated in the case of his-
tone 3-like centromeric protein A (CENP-A). Even
though this protein shows cleavage of the initiating meth-
oinine after nucleosome deposition, Bailey et al found
that approximately 10% of pre-nucleosomal CENP-A still
retains the initiating methionine [65]. Thus, the notation
of the AA-sequence changes for CENP-A depends on
its biological state. Since CENP-A H4 heterodimers form
a pre-nucleosomal complex with Holliday junction rec-
ognition protein (HJURP), which is active in G2 and
during mitosis, it is speculated that the variant with the
retained initiating methionine might be involved in
tumour development [65].

Due to the coexistence of the two reporting systems,
namely the legacy AA-based nomenclature of the
mature histone protein as well as the DNA-based
nomenclature, as used in genetic testing reports, there
are many ambiguous reports in the literature. However,
combined use of gene symbols, DNA sequence variants
(based on reference transcript sequences), and the pre-
dicted resulting substitutions of the protein unequiv-
cally defines the site of the sequence alteration.
Furthermore, it enables a DNA-based description of his-
tone sequence variants, which is regarded as the gold
standard for the reporting of variants according to the
HGVS sequence variant recommendations, and allows
unambiguous description of these alterations.

Application of the proposed nomenclature on
histone gene alterations found in brain tumours

The most common CNS tumour harbouring H3-3A
sequence variants is the ‘diffuse midline glioma, H3
K27-altered’. In this entity, a c.83A>T sequence variant
within H3-3A leads to the replacement of the AA lysine
(Lys) with methionine (Met) (p.Lys28Met). Such alter-
ations should be reported as H3-3A:c.83A>T p.Lys28-
Met (K27M) [15].

The second most frequent glial tumour harbouring an
H3-3A sequence variant is the diffuse hemispheric gli-
oma, H3 G34-mutant, which is characterised by the sub-
stitution of glycine (Gly) by either arginine (Arg) or
valine (Val) (p.Gly35Arg or p.Gly35Val). Following
the notation proposed here, the sequence alteration
should be reported as H3-3A:c.103G>C p.Gly35Arg
(G34R), or H3-3A:c.103G>A p.Gly35Arg
(G34R), or H3-3A:c.104G>T p.Gly35Val (G34V) [15].

The same reporting system would apply for the single
cases of gangliogliomas, pilocytic astrocytomas,
ependymomas [48], and subependymomas [51] that were reported to contain H3-3A sequence variants. It needs to be acknowledged that there is a debate about whether cases with histological features of a ganglioglioma and presence of a p.Gly35 substitution in the H3-3A-encoded protein are part of the spectrum of diffuse hemispheric glioma, previously reported as H3.3 G34-mutant, or whether they resemble real gangliogliomas with rare H3-3A sequence variation and are therefore not listed as such in COSMIC.

Application of the proposed nomenclature on histone gene alterations found in bone and soft tissue tumours

In the bone and soft tissue field, there are two bone tumour types that account for most of the soft tissue and bone tumours associated with recurrent histone sequence alterations: there is an established and consistent pattern of sequence variants in giant cell tumour of bone [23,26,35] with an H3-3A:c.103G>T p.Gly35Trp (G34W) substitution in the vast majority of those reported that supports the diagnosis. It should be noted that p.Lys37Met alterations predominantly encoded in H3-3B are seen in about 95% of chondroblastomas [23]. A few cases of chondrosarcoma, clear cell chondrosarcoma, and osteosarcoma [23,36] have also been reported to contain H3.3 histone A p.Lys37Met substitutions. The vast majority of chondroblastomas harbour p. Lys37Met encoded by H3-3B as opposed to H3-3A. Such sequence variants should be reported, for example, as H3-3A:c.110A>T p.Lys37Met (K36M) or H3-3B: c.110A>T p.Lys37Met (K36M).

Techniques for the detection of sequence alterations

Various techniques are available for analysing tissue samples for sequence variants using methods based on DNA sequencing or PCR. Some individual methods are capable of determining the number of variant DNA molecules relative to the normal reference sequence and hence are recognised as being quantitative or semi-quantitative. The choice of method and interpretation of results should depend upon the experience of the institution performing the analysis [66,67]. Also, the use of immunohistochemistry has become a routine in tumour diagnostics, as it is a fast and inexpensive method to report the presence or absence of tumour-related AA sequence alterations in neoplastic tissue. In addition, it also enables pathologists to clearly define the distribution and the cellular components that bear such sequence alterations. This not only allows for a better understanding of tumour types but also defines areas of tumour tissue where molecular analyses would be most promising for revealing conclusive results (by avoiding cellular components that are negative for these alterations). In addition to the sequence variant-specific antibodies against IDH1 R132H and BRAF V600E (BRAF V600E) that are commonly used in tumour diagnostics, there are also antibodies available that detect H3 isoform-

| Antibody | Suggested antibody name | Clone | Target | Host/isotype | Reactivity | Websites [Accessed 20 September 2020] |
|----------|-------------------------|-------|--------|-------------|------------|-------------------------------------|
| H3.3 K27M | H3.3 K28M | Recombinant monoclonal (RM192) | H3.3 p.K28M | Rabbit/IgG | Human | https://www.thermoscientific.com/antibody/product/H3-3-K27M-oncohistone-mutant-Antibody-clone-RM192-Recombinant-Monoclonal/MA5-27916 |
| H3K27me3 | H3 K28me3 | Monoclonal (G.299.10) | H3 K28me3 | Rabbit/IgG | Human, mouse, non-human primate, rat, Xenopus, zebrafish | https://www.thermoscientific.com/antibody/product/H3K27me3-Antibody-clone-G-299-10-Monoclonal/MA5-11198 |
| H3.3 G34R | H3.3 G35R | RM240 | H3.3 p.G35R | Rabbit/IgG | All | https://www.revmab.com/index.php/product/anti-histone-h3-3-g34r-rabbit-monoclonal-antibody-clone-rm240-histone-h3-3-g34r-mutant/ |
| H3.3 G34V | H3.3 G35V | 329E5; RM307 | H3.3 p.G35V | Rabbit/IgG | All | https://www.revmab.com/index.php/product/anti-histone-h3-3-g34v-rabbit-monoclonal-antibody-clone-rm307-histone-h3-3-g34v-mutant/ |
| H3.3 G34W | H3.3 G35W | RM263 | H3.3 p.G35W | Rabbit/IgG | All | https://www.revmab.com/index.php/product/anti-histone-h3-3-g34w-rabbit-monoclonal-antibody-clone-rm263-histone-h3-3-g34w-mutant/ |
| H3 K36M | H3 K37M | RM193 | H3.3 p.K37M | Rabbit/IgG | All | https://www.revmab.com/index.php/product/anti-histone-h3-k36m-rabbit-monoclonal-antibody-clone-rm193-histone-h3-k36m-mutant/ |
related alterations, such as H3.3 p.Lys28Met (K27M), H3.3 p.Gly35Arg (G34R) or H3.3 p.Gly35Val (G34V), as well as H3.3 p.Gly35Trp (G34W) and H3.3 p.Lys37Met (K36M) (Table 4).

Even though a positive immunohistochemical result is considered to be sufficient to define the tumour entity, further elucidation of a sequence variant’s prognostic value, e.g. in H3 K27-altered diffuse midline glioma located in the pons (DIPG), may depend on the specific gene in which the sequence is changed (see above), and hence requires additional molecular testing [59–61] (Table 5). Combined use of molecular analysis and immunohistochemistry also further controls for false-positive or false-negative results, which have been reported for both sequencing and immunohistochemistry [68]. Histone sequence variant-specific antibodies have been developed and named according to legacy histone AA numbering. To avoid confusion with the labelling of these antibodies, we suggest for the time being that these legacy descriptions be added in brackets following the HGVS-compliant standard notation (see above).

**Practical handling of the proposed new nomenclature**

Knowledge of the subject as summarised in this present paper has several important limitations. Ascertainment bias is a known issue in databases such as COSMIC, leading to a greater number of rare sequence variants being listed which then appear to be more common than they are. Reporting bias in publications is also likely, as there are case reports or small series rather than comprehensive studies, presenting an incomplete picture that requires more research. Sequencing of histone genes on a population basis in multiple geographical areas to account for any ethnic differences would be very helpful. Studies using tumour banks may fulfil this need, at least for the more common tumour types. Nevertheless, we have been able to identify some tumour types in which the reports are frequent enough to suggest that histone sequence variation may be important in their biology.

Since sequencing is usually more time-consuming than histomorphological and immunohistochemical assessment, the proposed nomenclature of entities with a histone sequence variant encourages a two-stage diagnosis. With the help of variant-specific antibodies and histone sequence variant-defining immunohistochemical features, a diagnosis can be given which includes the histomorphology and the suspected histone sequence variant based on the immunohistochemical findings, such as diffuse midline glioma with immunohistochemical detection of a p.Lys28Met (H3 K27M) sequence variant in one of the histone 3 isoforms. In some circumstances, such as in H3 K27-altered diffuse midline glioma located in the pons (DIPG), sequencing may not only be helpful to confirm the diagnosis but may also be of prognostic value, as indicated above. Following the proposed nomenclature, cases of diffuse hemispheric glioma, H3 G34-mutant, with, for example, immunohistochemical detection of H3.3 G34R positivity, might also be genetically evaluated to confirm changes in the nucleotide sequences, which have so far been described to be either H3-3A:c.103G>A p.Gly35Arg (G34R) or H3-3A:c.103G>C p.Gly35Arg (G34R).

**Conclusion**

This overview of tumour entities with sequence alterations of the H3-3A gene (HGNC ID = HGNC:4764) highlights the problems associated with the coexistence of genomic and traditional histological diagnostic tools.
of two distinct nomenclatures and proposes a standard for the reporting of histone sequence variants that allows an unambiguous description of these variants according to HGVS principles. Here, the gene, DNA, derived protein sequence, and legacy (antibody) sequence are given to account for the inconsistency in numbering resulting from the assumed post-translational modification of the protein. We hope that scientific journals will adopt the new notation, and that both geneticists and pathologists will find it helpful to include it in their reports.

Acknowledgements

We thank in particular Johan den Dunnen for the discussions on HGVS variant nomenclature and Elspeth Bruford for the discussions on gene nomenclature. In addition, we are grateful to the many other colleagues with whom we have discussed this matter and look forward to working with them to implement this proposal.

The content of this review represents the personal views of the authors and does not represent the views of the authors’ employers and associated institutions. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Author contributions statement

HL and IC conceived the manuscript based on discussions during publication of the 5th Edition of the WHO Classification of Tumours. HL and RD wrote the manuscript with input from IC, AL and GR.

References

1. den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat 2016; 37: 564–569.
2. Deans ZC, Fairley JA, den Dunnen JT, et al. HGVS nomenclature in practice: an example from the United Kingdom National External Quality Assessment Scheme. Hum Mutat 2016; 37: 576–578.
3. Ogino S, Gulley ML, den Dunnen JT, et al. Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. J Mol Diagn 2007; 9: 1–6.
4. den Dunnen JT, Paalman MH. Standardizing mutation nomenclature: why bother? Hum Mutat 2003; 22: 181–182.
5. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015; 17: 405–424.
6. Li MM, Datto M, Duncavage EJ, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn 2017; 19: 4–23.
7. Maze I, Noh KM, Soshnev AA, et al. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat Rev Genet 2014; 15: 259–271.
8. Scott WA, Campos EI. Interactions with histone H3 & tools to study them. Front Cell Dev Biol 2020; 8: 701.
9. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol 2015; 16: 178–189.
10. Urahama T, Harada A, Maebara K, et al. Histone H3.5 forms an unstable nucleosome and accumulates around transcription start sites in human testis. Epigenetics Chromatin 2016; 9: 2.
11. DeLange RJ, Hooper JA, Smith EL. Complete amino-acid sequence of calf-thymus histone 3. Proc Natl Acad Sci U S A 1972; 69: 882–884.
12. DeLange RJ, Famborough DM, Smith EL, et al. Calf and pea histone IV. 3. Complete amino acid sequence of pea seedling histone IV; comparison with the homologous calf thymus histone. J Biol Chem 1969; 244: 5669–5679.
13. Iwai K, Ishikawa K, Hayashi H. Amino-acid sequence of slightly lysine-rich histone. Nature 1970; 226: 1056–1058.
14. Turner BM. Reading signals on the nucleosome with a new nomenclature for modified histones. Nat Struct Mol Biol 2005; 12: 110–112.
15. Leske H, Rushing E, Budka H, et al. K27/G34 versus K28/G35 in histone H3-mutant gliomas: a note of caution. Acta Neuropathol 2018; 136: 175–176.
16. Sayers EW, Beck J, Brister JR, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2020; 48: D9–D16.
17. Crea IA, Deans Z, Ligtenberg MJ, et al. Guidance for laboratories performing molecular pathology for cancer patients. J Clin Pathol 2014; 67: 923–931.
18. Deans ZC, Costa JL, Cree I, et al. Integration of next-generation sequencing in clinical diagnostic molecular pathology laboratories for analysis of solid tumours; an expert opinion on behalf of IQN Path ASBL. Virchows Arch 2017; 470: 5–20.
19. Bruford EA, Braschi B, Denny P, et al. Guidelines for human gene nomenclature. Nat Genet 2020; 52: 754–758.
20. Freeman PJ, Hart RK, Grettôn JJ, et al. VariantValidator: accurate validation, mapping, and formatting of sequence variation descriptions. Hum Mutat 2018; 39: 61–68.
21. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009; 6: e1000097.
22. Toledo RA, Qin Y, Cheng ZM, et al. Recurrent mutations of chromatin-remodeling genes and kinase receptors in pheochromocytomas and paragangliomas. Clin Cancer Res 2016; 22: 2301–2310.
23. Behjati S, Tarpey PS, Preseanu N, et al. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat Genet 2013; 45: 1479–1482.
24. Castel D, Philippe C, Calmon R, et al. Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. Acta Neuropathol 2015; 130: 813–827.
25. Schwartzentruber J, Korshunov A, Liu XY, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature 2012; 482: 226–231.
26. Rigli A, Mancini I, Gambardotti M, et al. Histone 3.3 mutations in giant cell tumor and giant cell-rich sarcomas of bone. Hum Pathol 2017; 68: 128–135.
27. Mackay A, Burford A, Molinari V, et al. Molecular, pathological, radiological, and immune profiling of non-brainstem pediatric high-grade glioma from the HERBY phase II randomized trial. Cancer Cell 2018; 33: 829–842.e5.
28. Yoshida KJ, Nakano Y, Honda-Kikuhara M, et al. Absence of H3F3A mutation in a subset of malignant giant cell tumor of bone. Mod Pathol 2019; 32: 1751–1761.
A classification framework for histone alterations

29. Mackay A, Burford A, Carvalho D, et al. Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma. Cancer Cell 2017; 32: 520–537.e5.

30. Yoshimoto K, Hatae R, Sagasuda Y, et al. Prevalence and clinicopathological features of H3.3 G34-mutant high-grade gliomas: a retrospective study of 411 consecutive glioma cases in a single institution. Brain Tumor Pathol 2017; 34: 103–112.

31. Pathak P, Jha P, Purkait S, et al. Altered global histone-trimethylation code and H3F3A-ATRX mutation in pediatric GBM. J Neurooncol 2015; 121: 489–497.

32. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017; 23: 703–713.

33. Martin D, Abba MC, Molinolo AA, et al. The head and neck cancer oncogene: a platform for the development of precision molecular therapies. Oncotarget 2014; 5: 8906–8923.

34. Nacev BA, Feng L, Bagert JD, et al. The expanding landscape of ‘onco-histone’ mutations in human cancers. Nature 2019; 567: 473–478.

35. Lim J, Park JH, Baude A, et al. The histone variant H3.3 G34W substitution in giant cell tumor of the bone link chromatin and RNA processing. Sci Rep 2017; 7: 13459.

36. Koelsche C, Schrimpf D, Tharun L, et al. H3 K27M mutations are extremely rare in posterior fossa group A ependymoma. Brain Pathol 2017; 27: 1682–1689.

37. Pfafflin-Cavanagh S, Lu C, Gayden T, et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat Genet 2017; 49: 180–185.

38. Lohr JG, Stojanov P, Lawrence MS, 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 2012; 109: 3879–3884.

39. Huether R, Dong L, Chen X, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. Nat Commun 2014; 5: 3630.

40. Wagner AH, Devarakonda S, Skidmore ZL, et al. Immunohistochemistry for H3K27M mutation in T-cell acute lymphoblastic leukemia. Br J Haematol 2019; 184: 676–679.

41. Collord G, Martincorena I, Young MD, et al. Recurrent histone mutations in T-cell acute lymphoblastic leukemia. Br J Haematol 2017; 174: 40–51.
SUPPLEMENTARY MATERIAL ONLINE

**Figure S1.** Alignment of the coding sequence (CDS) of the reference transcript sequences NM_002107.7 and NM_005324.5 for the histone genes H3-3A and H3-3B, respectively, with the respective identical amino acid reference sequences NP_002098.1 and NP_005315.1 of histone protein H3.3

**Table S1.** Tabulated results of tumours listed in COSMIC for H3F3A sequence variants