Tag	Type	Description
H1	i	Number of 1-difference hits (see also NM)
H2	i	Number of 2-difference hits
HI	i	Query hit index
IH	i	Query hit total count
LB	Z	Library
MC	Z	CIGAR string for mate/next segment
MD	Z	String encoding mismatched and deleted reference bases
MF	?	Reserved for backwards compatibility reasons
MI	Z	Molecular identifier; a string that uniquely identifies the molecule from which the record was derived
ML	B,C	Base modification probabilities
MM	Z	Base modifications / methylation
MQ	i	Mapping quality of the mate/next segment
MZ	i	Length of sequence at the time MM and ML were produced
NH	i	Number of reported alignments that contain the query in the current record
NM	i	Edit distance to the reference
OA	Z	Original alignment
OC	Z	Original CIGAR (deprecated; use OA instead)
OP	i	Original mapping position (deprecated; use OA instead)
OQ	Z	Original base quality
OX	Z	Original unique molecular barcode bases
PG	Z	Program
PQ	i	Phred likelihood of the template
PT	Z	Read annotations for parts of the padded read sequence
PU	Z	Platform unit
Q2	Z	Phred quality of the mate/next segment sequence in the R2 tag
QT	Z	Phred quality of the sample barcode sequence in the BC tag
QX	Z	Quality score of the unique molecular identifier in the RX tag
R2	Z	Sequence of the mate/next segment in the template
RG	Z	Read group
RT	?	Reserved for backwards compatibility reasons
RX	Z	Sequence bases of the (possibly corrected) unique molecular identifier
S2	?	Reserved for backwards compatibility reasons
SA	Z	Other canonical alignments in a chimeric alignment
SM	i	Template-independent mapping quality
SQ	?	Reserved for backwards compatibility reasons
TC	i	The number of segments in the template
TS	A	Transcript strand
U2	Z	Phred probability of the 2nd call being wrong conditional on the best being wrong
UQ	i	Phred likelihood of the segment, conditional on the mapping being correct
X?	?	Reserved for end users
Y?	?	Reserved for end users
Z?	?	Reserved for end users

1.1 Additional Template and Mapping data

AM:i:`score` The smallest template-independent mapping quality of any segment in the same template as this read. (See also `SM`.)

AS:i:`score` Alignment score generated by aligner.

BQ:Z:`qualities` Offset to base alignment quality (BAQ), of the same length as the read sequence. At the i-th read base, $\mathrm{BAQ}_i = Q_i - (\mathrm{BQ}_i - 64)$ where Q_i is the i-th base quality.

CC:Z:`rname` Reference name of the next hit; '`=`' for the same chromosome.

Unmodified base	Code	Abbreviation	Name	ChEBI
C	m	5mC	5-Methylcytosine	27551
C	h	5hmC	5-Hydroxymethylcytosine	76792
C	f	5fC	5-Formylcytosine	76794
C	c	5caC	5-Carboxylcytosine	76793
C	C		Ambiguity code; any C mod	
T	g	5hmU	5-Hydroxymethyluracil	16964
T	e	5fU	5-Formyluracil	80961
T	b	5caU	5-Carboxyluracil	17477
T	T		Ambiguity code; any T mod	
U	U		Ambiguity code; any U mod	
A	a	6mA	6-Methyladenine	28871
A	A		Ambiguity code; any A mod	
G	o	8oxoG	8-Oxoguanine	44605
G	G		Ambiguity code; any G mod	
N	n	Xao	Xanthosine	18107
N	N		Ambiguity code; any mod	

ML:B:C,`scaled-probabilities`

The optional `ML` tag lists the probability of each modification listed in the `MM` tag being correct, in the order that they occur. The continuous probability range 0.0 to 1.0 is remapped in equal sized portions to the discrete integers 0 to 255 inclusively. Thus the probability range corresponding to integer value N is $N/256$ to $(N+1)/256$.

The SAM encoding therefore uses a byte array of type '`C`' with the number of elements matching the summation of the number of modifications listed as being present in the `MM` tag accounting for multi-modifications each having their own probability.

For example '`MM:Z:C+m,5,12;C+h,5,12;`' may have an associated tag of '`ML:B:C,204,89,26,130`'.

If the above is rewritten in the multiple-modification form, the probabilities are interleaved in the order presented, giving '`MM:Z:C+mh,5,12; ML:B:C,204,26,89,130`'. Note where several possible modifications are presented at the same site, the `ML` values represent the absolute probabilities of the modification call being correct and not the relative likelihood between the alternatives. These probabilities should not sum to above 1.0 ($\approx$ 256 in integer encoding, allowing for some minor rounding errors), but may sum to a lower total with the remainder representing the probability that none of the listed modification types are present. In the example used above, the 6th `C` has 80% chance of being `5mC`, 10% chance of being `5hmC` and 10% chance of being an unmodified `C`.

`ML` values for ambiguity codes give the probability that the modification is one of the possible codes compatible with that ambiguity code. For example `MM:Z:C+C,10; ML:B:C,229` indicates a C call with a probability of 90% of having some form of unspecified modification.

MZ:i:`length`

Tools may edit the SEQ sequence data, such as modifying the alignment with hard-clipping. If the sequence is shrunk in this manner then the base offsets in `MM` and `ML` become invalid unless they are also updated accordingly.

There may be hard-clipping tools which update `MM` and tools which do not, so the `MZ` tag offers a simple sanity check. It holds the length of the sequence at the time `MM` was last written. Tools that wish to validate `MM` should compare the length of the SEQ field with the contents of the `MZ` tag. The tag is optional, but recommended, and if it is absent then there is an implicit assumption that the `MM` data is valid unless evidence implies otherwise (such as having coordinates beyond the end of the sequence).

2 Draft tags

These are tags which have been proposed and are broadly accepted to become standard tags, but a review or probationary period has been deemed useful. They use the locally-defined tag namespace and processing