

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 software should consider that the tags may have local usage for other purposes.

~~There are currently no tags with draft status~~

<u>Tag</u>	<u>Type</u>	<u>Description</u>
<u>An</u>	<u>Z</u>	<u>Molecular annotation names</u>
<u>Aq</u>	<u>B,C</u>	<u>Molecular annotation qualities</u>
<u>Ma</u>	<u>Z</u>	<u>Molecular annotations</u>

2.1 Molecular annotations

The molecular annotation tags (Ma, Aq, and An) describe intervals (annotations) on the sequence as originally reported by the sequencing instrument. These annotations describe molecular features (annotation types) spanning one or more bases, for example nucleosome positions, methylation-accessible patches, and protein binding sites.

Molecular annotations use three related tags. Ma records the annotation types and annotations. Aq optionally records per annotation quality values. An optionally records per annotation names.

As with base modifications, positions are recorded against the original orientation of the sequenced molecule. Starts in Ma are 1-based.

Ma:Z:length(;type[-+.][PQ]*:start-length(,start-length)*)+

The first field is the length of SEQ at the time the Ma value was last written. This allows tools to detect when SEQ has later been trimmed, clipped, or otherwise altered. If this length differs from the current SEQ length, the interval positions in the Ma tag should be considered invalid.

Each block begins with an annotation type name, followed by a strand character and optional quality specifier. An annotation type is the feature being reported, such as a nucleosome, methylation-accessible patch, or protein-binding site. It is analogous to a base modification code in MM. The intervals in the block are individual annotations of that type. For example:

Ma:Z:1000;ctcf+Q:100-25,700-30;ctcf-Q:800-20

Here ctcf is the annotation type. The tag contains three annotations: two on the '+' strand and one on the '-' strand. Both blocks describe the same annotation type.

Type names contain letters, digits, or underscores and are case-sensitive. Strand is one of '+' for the same strand as the original sequenced molecule, '-' for the opposite strand, or '.' for unstranded, unknown, or inapplicable strand.

The optional quality specifier declares the number and interpretation of Aq values per annotation. 'P' indicates one Phred-scaled confidence value. 'Q' indicates one linearly scaled probability value, using the same byte mapping as ML. Multiple characters, such as 'PQ', indicate multiple Aq values per annotation in the stated order. If the specifier is absent, no Aq values are stored for annotations of that annotation type.

The interval list is a comma-separated list of intervals written start-length. Within each block, intervals must be listed in non-decreasing start order. Lengths are the number of bases spanned by the annotation, so 101-50 covers bases 101 through 150 inclusive. Starts and lengths must be positive,

and no interval may extend beyond the length recorded at the start of `Ma`. Individual `annotations` of the same `annotation type` may overlap.

If a type name is used in more than one block, all such blocks must use the same quality specifier. Parsers should reject an `Ma` tag where one type name is used with conflicting quality specifiers.

`Aq:B:C`, qualities

The optional `Aq` tag lists quality values for `annotations` in `Ma` order, stored as a `B,C` array of unsigned 8-bit integer (`uint8_t`) values. It is omitted when no `annotation type` has quality values. The number of `Aq` elements must equal the sum, over all `Ma` blocks, of the number of `annotations` multiplied by the number of quality-specifier characters in that block.

`P` values use the Phred-scaled confidence convention defined by the main SAM specification. `Q` values use the same 0-255 linearly scaled probability encoding as `ML`. For example:

```
Ma:Z:1000;msp.P:100-50,200-60;nuc.:500-147
Aq:B:C,40,35,
```

This gives Phred qualities 40 and 35 to the two `msp` `annotations`; the `nuc` `annotation` has no quality value.

`An:Z`:names

The optional `An` tag lists comma-separated names or labels for individual `annotations` in `Ma` order. `An` names `annotations`, not `annotation types`. If no `annotations` have names, `An` should be omitted.

If only some `annotations` are named, empty fields are used for unnamed entries. For example:

```
Ma:Z:1000;msp.P:100-50,200-60,320-45;nuc.:500-147
An:Z:msp1,,msp3,,
```

The first and third `msp` `annotations` are named, while the second `msp` and following `nuc` `annotation` are unnamed. These names do not define new `annotation types`. The number of fields in `An` must equal the total number of intervals in `Ma`. Names must not contain commas.

3 Locally-defined tags

You can freely add new tags. Note that tags starting with ‘X’, ‘Y’, or ‘Z’ and tags containing lowercase letters in either position are reserved for local use and will not be formally defined in any future version of this specification.

If a new tag may be of general interest, it may be useful to have it added to this specification. Additions can be proposed by opening a new issue at <https://github.com/samtools/hts-specs/issues> and/or by sending email to samtools-devel@lists.sourceforge.net.

Appendix A Tag History

This appendix lists when standard tags were initially defined or significantly changed, and other historical events that affect how tags are interpreted or what files they may appear in.

July 2026

Added the `Ma`, `Aq`, and `An` draft tags describing molecular annotations.

September 2024

Added the `MN` tag for validating base modification tag consistency.