PT:Z:*annotag*`(\|`*annotag*`)*` where each *annotag* matches *start*`;`*end*`;`*strand*`;`*type*`(;`*key*`(=`*value*`)?)*`
Read annotations for parts of the padded read sequence.

The `PT` tag value has the format of a series of annotation tags separated by '`|`', each annotating a sub-region of the read. Each tag consists of *start*, *end*, *strand*, *type* and zero or more *key*=*value* pairs, each separated with semicolons. *Start* and *end* are 1-based positions between one and the sum of the `M/I/D/P/S/=/X` CIGAR operators, i.e., SEQ length plus any pads. Note any editing of the CIGAR string may require updating the `PT` tag coordinates, or even invalidate them. As in GFF3, *strand* is one of '`+`' for forward strand tags, '`-`' for reverse strand, '`.`' for unstranded or '`?`' for stranded but unknown strand.

The *type* and any *keys* and their optional *values* are all percent encoded as in the `CT` tag.

1.6 Technology-specific data

FZ:B:S,`intensities` Flow signal intensities on the original strand of the read, stored as `(uint16_t) round(value * 100.0)`.

1.6.1 Color space

CM:i:`distance` Edit distance between the color sequence and the color reference (see also `NM`).

CS:Z:`sequence` Color read sequence on the original strand of the read. The primer base must be included.

CQ:Z:`qualities` Color read quality on the original strand of the read. Same encoding as QUAL; same length as `CS`.

1.7 Base modifications

Base modifications, including base methylation, are represented as a series of edits from the primary unmodified sequence as originally reported by the sequencing instrument. This potentially differs to the sequence stored in the main SAM SEQ field if the latter has been reverse complemented, in which case SAM FLAG 0x10 must be set. This means modification positions are also recorded against the original orientation (i.e. starting at the 5' end), and count the original base types.

Each modified base prediction listed also has a quality value associated with it. Given the unmodified base already has a phred likelihood, this base modification quality should be interpreted as the likelihood of this modification being correct given an assumption the original call is correct.

~~MM:Z:~~

MM:Z:`([ACGTUN][-+]([A-Za-z]+|[0-9]+)[.?]?(,[0-9]+)*;)*`
The first character is the unmodified "fundamental" base as reported by the sequencing instrument for the top strand. It must be one of '`A`', '`C`', '`G`', '`T`', '`U`' (if RNA) or '`N`' for anything else, including any IUPAC ambiguity codes in the reported SEQ field. Note '`N`' may be used to match any base rather than specifically an '`N`' call by the sequencing instrument. This may be used in situations where the base modification is not a derivation of a standard base type. This is followed by either plus or minus indicating the strand the modification was observed on (relative to the original sequenced strand of SEQ with plus meaning same orientation),[3] and one or more base modification codes.

Following the base modification codes is a recommended but optional '`.`' or '`?`' describing how skipped seq bases of the stated base type should be interpreted by downstream tools. When this flag is '`?`' there is no information about the modification status of the skipped bases provided. When this flag is not present, or it is '`.`', these bases should be assumed to have low probability of modification.[4]

[3] Hence a tool that may reverse complement sequences does not need to understand how to manipulate the `MM` and `ML` tags.

[4] The decision whether a base is assumed to be unmodified or has a probability explicitly provided is up to the modification calling program. Some programs will elide calls with modification probabilities below a threshold to provide a more compact modification tag.