

2 Recommended Practice for the SAM Format

This section describes the best practice for representing data in the SAM format. They are not required in general, but may be required by a specific software package for it to function properly.

1. The header section
 - 1 The `@HD` line should be present, with either the `SO` tag or the `GO` tag (but not both) specified.
 - 2 The `@SQ` lines should be present if reads have been mapped.
 - 3 When a `RG` tag appears anywhere in the alignment section, there should be a single corresponding `@RG` line with matching ID tag in the header.
 - 4 When a `PG` tag appears anywhere in the alignment section, there should be a single corresponding `@PG` line with matching ID tag in the header.
2. Adjacent CIGAR operations should be different.
3. No alignments should be assigned mapping quality 255.
4. Unmapped reads
 - 1 For a unmapped paired-end or mate-pair read whose mate is mapped, the unmapped read should have `RNAME` and `POS` identical to its mate.
 - 2 If all segments in a template are unmapped, their `RNAME` should be set as `*` and `POS` as 0.
 - 3 If `POS` plus the sum of lengths of `M/=X/D/N` operations in CIGAR exceeds the length specified in the `LN` field of the `@SQ` header line (if exists) with an `SN` equal to `RNAME`, the alignment should be unmapped, unless the reference sequence is circular (see below).
 - 4 Unmapped reads should be stored in the orientation in which they came off the sequencing machine and have their `reverse` flag bit (0x10) correspondingly unset.
 - 5 [Unmapped reads should have mapping quality zero.](#)
5. Multiple mapping
 - 1 When one segment is present in multiple lines to represent a multiple mapping of the segment, only one of these records should have the secondary alignment flag bit (0x100) unset. `RNEXT` and `PNEXT` point to the primary line of the next read in the template.
 - 2 `SEQ` and `QUAL` of secondary alignments should be set to `*` to reduce the file size.
6. Optional tags:
 - 1 If the template has more than 2 segments, the `TC` tag should be present.
 - 2 The `NM` tag should be present.

7. Circular reference sequences

Mappings that cross the coordinate ‘join’ in circular reference sequences (i.e., those whose `@SQ` headers specify `TP:circular`) may be represented as follows:

- 1 (Preferred) As usual `POS` should be between 1 and the `@SQ` header’s `LN` value, but `POS` plus the sum of the lengths of `M/=X/D/N` CIGAR operations may exceed `LN`. Coordinates greater than `LN` are interpreted by subtracting `LN` so that bases at `LN+1`, `LN+2`, `LN+3`, ... are considered to be mapped at positions 1, 2, 3, ...; thus each (1-based) position p is interpreted as $((p - 1) \bmod LN) + 1$.²⁰
- 2 Alternatively, such alignments may be split across several records: one record representing the initial portion of the segment ending at `LN`, one representing the final portion starting from 1, and any other records representing additional portions in between spanning the entire reference sequence. One record (chosen arbitrarily) is considered primary and the remainder have their `supplementary` flag bit (0x800) set.

²⁰The impact of this representation on indexing and random access is yet to be explored by implementations.