

DT	Date the run was produced (ISO8601 date or date/time).
FO	Flow order. The array of nucleotide bases that correspond to the nucleotides used for each flow of each read. Multi-base flows are encoded in IUPAC format, and non-nucleotide flows by various other characters. <i>Format:</i> /\*  [ACMGRSVTWYHKDBN] +/
KS	The array of nucleotide bases that correspond to the key sequence of each read.
LB	Library.
PG	Programs used for processing the read group.
PI	Predicted median insert size, rounded to the nearest integer.
PL	Platform/technology used to produce the reads. <i>Valid values:</i> CAPILLARY, DNBSEQ (MGI/BGI), ELEMENT, HELICOS, ILLUMINA, IONTORRENT, LS454, ONT (Oxford Nanopore), PACBIO (Pacific Biosciences), SINGULAR, SOLID, and ULTIMA. <sup>11</sup> This field should be omitted when the technology is not in this list (though the PM field may still be present in this case) or is unknown.
PM	Platform model. Free-form text providing further details of the platform/technology used.
PU	Platform unit (e.g., flowcell-barcode.lane for Illumina or slide for SOLiD). Unique identifier.
SM	Sample. Use pool name where a pool is being sequenced.
@PG	Program.
ID*	Program record identifier. Each @PG line must have a unique ID. The value of ID is used in the alignment PG tag and PP tags of other @PG lines. PG IDs may be modified when merging SAM files in order to handle collisions.
PN	Program name
CL	Command line. UTF-8 encoding may be used.
PP	Previous @PG-ID. Must match another @PG header's ID tag. @PG records may be chained using PP tag, with the last record in the chain having no PP tag. This chain defines the order of programs that have been applied to the alignment. PP values may be modified when merging SAM files in order to handle collisions of PG IDs. The first PG record in a chain (i.e., the one referred to by the PG tag in a SAM record) describes the most recent program that operated on the SAM record. The next PG record in the chain describes the next most recent program that operated on the SAM record. The PG ID on a SAM record is not required to refer to the newest PG record in a chain. It may refer to any PG record in a chain, implying that the SAM record has been operated on by the program in that PG record, and the program(s) referred to via the PP tag.
DS	Description. UTF-8 encoding may be used.
VN	Program version
@CO	One-line text comment. Unordered multiple @CO lines are allowed. UTF-8 encoding may be used.

### 1.3.1 Defined sub-sort terms

While the **SS** sub-sort field allows implementation-defined keywords, some terms are predefined with specific meanings.

**lexicographical** sort order is defined as a character-based dictionary sort with the character order as defined by the POSIX C locale. For example “abc”, “abc17”, “abc5”, “abc59” and “abcd” are in lexicographical order.

**natural** sort order is similar to lexicographical order except that runs of adjacent digits are considered to be numbers embedded within the text string, ordered numerically when compared to each other and ordered as single digits when compared to the surrounding non-digit characters. Runs that differ only in the number of leading zeros (thus are numerically tied) are ordered by more-zeros coming before fewer-zeros. The characters ‘-’ and ‘.’ are considered as ordinary characters, so apparently negative or fractional values are not treated as part of an embedded number. For example, “abc”, “abc+5”, “abc-5”, “abc.d”, “abc03”, “abc5”, “abc008”, “abc08”, “abc8”, “abc17”, “abc17.+”, “abc17.2”, “abc17.d”, “abc59” and “abcd” are in natural order.

<sup>11</sup>[The PL value should be written in uppercase exactly as shown in this list of valid values. Tools should also accept lowercase when reading the @RG PL field, due to the existence of public data files with lowercase PL values.](#)