

## SAMtools View Output Explained

The samtools view command is used to view, filter, and convert alignment files in SAM, BAM, or CRAM formats.

### Output Format (SAM/BAM)

Each line in a SAM file represents a single read and contains several fields:

Field	Description
QNAME	Query name of the read
FLAG	Bitwise flag indicating read properties (e.g., paired, mapped)
RNAME	Reference sequence name
POS	1-based leftmost mapping position
MAPQ	Mapping quality
CIGAR	CIGAR string showing alignment
RNEXT	Reference name of the mate/next read
PNEXT	Position of the mate/next read
TLEN	Template length
SEQ	Read sequence
QUAL	Base quality scores

Additional optional fields may follow, providing tags with extra information like alignment scores, read groups, or platform-specific data.

#### Common Use Cases

Convert BAM to SAM: `samtools view -h input.bam > output.sam`

Filter reads by mapping quality: `samtools view -q 30 input.bam`

Extract reads mapped to a specific region: `samtools view input.bam chr1:1000-2000`

## ■ Standard BED Format (3–12 Columns)

Column	Name	Description
1	<code>chrom</code>	Chromosome name (e.g., <code>chr1</code> , <code>chrX</code> )
2	<code>start</code>	Start position (0-based)
3	<code>end</code>	End position (exclusive)
4	<code>name</code>	Optional name of the feature
5	<code>score</code>	Optional score (0–1000)
6	<code>strand</code>	Optional strand ( <code>+</code> or <code>-</code> )
7–12	Additional fields for BED12 (e.g., block sizes, exon counts)	

## □ What `sed` Outputs

- **Line-by-line text:** `sed` processes input one line at a time.
- **Modified content:** It applies your editing instructions (like substitutions, deletions, insertions).
- **Printed to standard output:** Unless you redirect it to a file or suppress output with `-n`.

## Standard SQL Output Format

When you run a SQL query like SELECT, the output is usually presented as a table with rows and columns:

- Rows represent individual records.
- Columns represent fields selected in the query.
- Data types can include text, numbers, dates, booleans, et

## dplyr (R)

- Output Format: tibble (a modern version of data frames)
- Details:
  - Results from summarise(), mutate(), filter(), etc., are returned as tibbles.
  - Tibbles print in a compact, readable format with column types shown.
  - Can be exported to CSV, Excel, or other formats using write.csv(), write\_excel\_csv(), etc.

## awk

- Output Format: Plain text
- Details:
  - Uses print and printf for output.
  - print outputs space-separated values with newline.
  - printf allows formatted output (e.g., aligned columns, decimal precision).

## Apache Spark

- Output Format:
  - Batch: DataFrame (can be saved as Parquet, JSON, CSV, ORC, etc.)
  - Streaming: Console, Kafka, Parquet, Memory, etc.
- Details:
  - write.format("parquet").save("path") for batch
  - writeStream.format("console").outputMode("append").start() >>> for streaming

## Picard Tools

- Output Format: SAM, BAM, CRAM, VCF, FASTQ (depending on the tool)
- Details:
  - Tools like SamFormatConverter, AddOrReplaceReadGroups, CollectRnaSeqMetrics produce structured genomic data files.
  - Output format is determined by file extension or command-line flags.

## Apache (General)

- Output Format: Varies by project
  - Apache FOP: PDF, PS, PCL, AFP, XML, PNG, RTF, TXT
  - Apache FreeMarker: HTML, XML, JSON, plain text
- Details:
  - Output format depends on the renderer or template engine used.
  - Often used for web, document, or data serialization tasks.

## pandas (Python)

- Output Format: DataFrame
- Details:
  - Displays as aligned tables in console or styled HTML in Jupyter.
  - Can export to CSV, Excel, JSON, HTML, LaTeX, Markdown, etc.