Prerequisite Terminologies:

In order to have a better understanding of the main topic, you should have the basic concept of the following terms:

- ➢ **Next-generation Sequencing Data**
- ➢ **Reference Genome**

Introduction:

The SAM (Sequence Alignment Map) Format is a text format for storing sequence data in a series of tab delimited ASCII columns. Most often it is generated as a human readable version of its sister BAM format. Most SAM format data is output from aligners that read FASTQ files and assign the sequences to a position with respect to a known reference genome. In the future, SAM will also be used to archive unaligned sequence data generated directly from sequencing machines.

Format:

SAM file format contains an optional header section followed by the alignment section:

Header Section:

- The header section may contain information about the entire file and additional information about the alignments. The alignments then associate themselves with specific header information.
- The header section must be prior to the alignment section if it is present.
- Headings begin with the '@' symbol, which distinguishes them from the alignment section.

Alignment Section:

- The alignment section contains the information for each sequence about where/how it aligns to the reference genome.
- The alignment sections have 11 mandatory fields, as well as a variable number of optional fields.
- The fields are discussed in the following table:

Col	Field	Type	Range	Brief Description
1	QNAME	String	[!-?A - ~] {1,254}	Query template name
2	FLAG	Int	[0, 2^{16} - 1]	Bitwise FLAG
3	RNAME	String	*| [:rname:^*=][:rname:]*	Reference Sequence name
4	POS	Int	[0, 2^{31}- 1]	1- based leftmost mapping position
5	MAPQ	Int	[0, 2^{8}- 1]	Mapping quality
6	CIGAR	String	*|([0-9][MIDNSHPX=])+	CIGAR string
7	RNEXT	String	*|=| [:rname:^*=][:rname:]*	Ref. name of the next read
8	PNEXT	Int	[0, 2^{31}- 1]	Position of the next

				read
9	TLEN	Int	$[-2^{31}+1, [2^{31}-1]]$	Observed template length
10	SEQ	String	*|[A-Za-z=.]+	Segment sequence
11	QUAL	String	[! - ~]+	ASCII of Phred-scaled base quality +33

Practical Uses:

Following are the uses of SAM file format:

- It stores the next generation sequencing data.
- It contains sequence reads alignments. Both short and long reads are supported in SAM files upto 128Mbb.
- SAM file format is used in Genomics Analysis Toolkit (GATK), which contains many tools for genomic analysis.
- SAM file format also stores quality scores for the alignments.
- The sequencing data stored in SAM files are generated from CHIP-seq, RNA-seq and various other methodologies.

Constitutes of SAM:

There are various alignments stored in a single file of SAM format, such as:

- **Clipped Alignment:**
  In clipped alignment, sequencing reads allows the masking of portions of the reads that do not align to the genome from end to end, (sub-sequences at the ends may be clipped off), which may be desirable for certain types of analysis.
- **Spliced Alignment:**
  A spliced alignment is an alignment of a partial or full length spliced. transcript sequence against an unspliced genomic sequence. A spliced. alignment allows to highlight the boundaries and the alignment of exons of. the transcript sequence on the genomic sequence.

- **Multi-part Alignment:**
  In the Multi-part alignment, a single query sequence may be aligned to multiple parts of the reference genome, either with or without the overlapping.
- **Padded Alignment:**
  Most sequence aligners only give the sequences inserted to the reference genome, but do not present how these inserted sequences are aligned against each other. Alignment with inserted sequences fully aligned is called padded alignment. Padded alignment is always produced by de novo assemblers and is important for an alignment viewer to display the alignment properly.

File Extension:

The file extension for SAM file is ***.sam***.

Summary:

In this video tutorial of file formats, we have learned about the SAM file format. We also got to know SAM format, information stored in the header and alignment section, SAM format uses and different alignments stored within a single SAM file.