

Glossary of deep sequencing jargon

BAM	• binary file format (complement to SAM); should be the result of an alignment of sequenced reads to a reference genome • each line = 1 mapped read, with information about its mapping quality, its sequence, its location in the genome etc. • highly recommended format for storing data • to make a BAM file human-readable, one can, for example, use the program <code>samtools view</code>
bed	• text file • used for genomic intervals, e.g. genes, peak regions etc. • actually, there is a rather strict definition of the format that can be found at UCSC (http://genome.ucsc.edu/FAQ/FAQformat.html#format1) • for deepTools, the first 3 columns are important: chromosome, start position of the region, end position of the genome
bedGraph	• text file • similar to bed file, except that it can ONLY contain 4 columns and 4th column MUST be a score
bigWig	• binary version of a bedGraph file • usually contains 4 columns: chromosome, start of genomic bin, end of genomic bin, score • the score can be anything, e.g. an average read coverage • very useful to
FASTQ	• text file • contains raw read information (e.g. base calls, sequencing quality measures etc.), but not information about where in the genome the read originated from
SAM	• should be the result of an alignment of sequenced reads to a reference genome • each line = 1 mapped read, with information about its mapping quality, its sequence, its location in the genome etc. • it is recommended to generate the binary (compressed) version of this file format: BAM