**Common File Formats Used by the TaRGET II Consortium**

**Overview**

The TaRGET II consortium uses several file formats to store, display, and disseminate data:

* FASTQ: a text-based format for storing nucleotide sequences (reads) and their quality scores.
* BAM: The Sequence Alignment/Mapping (SAM) format is a text-based format for storing read alignments against reference sequences and it is interconvertible with the binary BAM format.
* bigWig: The bigWig format is an indexed binary format for rapid display of continuous and dense data.
* Stranded bedGraph (.sbg): The bedGraph format allows display of continuous-valued data in track format. This display type is useful for probability scores and transcriptome data. Stranded bedGraph separates RNA-seq reads that were mapped to the positive and the negative strands.
* narrowPeak: A narrowPeak (.narrowPeak) file is used to provide called peaks of signal enrichment for ATAC-seq

These file formats were originally designed to be generic and flexible. The consortium considers FASTQ as the basic file format for archival purpose and thus the FASTQ format's specifications aim to preserve the raw sequence data. In comparison, the other file formats are geared towards data visualization and dissemination, thus their specifications aim to facilitate user-friendliness.