

File formats CheatSheet

Format	Full Form	Description	Common Use
FASTA (.fasta, .fa)	Fast-All	Stores nucleotide or protein sequences	Sequence storage and retrieval
FASTQ (.fastq, .fq)	FastQ format	Stores sequencing reads with quality scores	Raw sequencing data from NGS
SAM (.sam)	Sequence Alignment Map	Text-based format for storing aligned sequencing reads	Sequence alignment against reference genome
BAM (.bam)	Binary Alignment Map	Binary version of SAM for efficient storage	Compressed sequence alignment data
CRAM (.cram)	Compressed Reference-Aligned Map	More compressed version of BAM, reducing storage space	Efficient storage and retrieval of sequencing alignments
VCF (.vcf)	Variant Call Format	Stores genetic variations (SNPs, indels, etc.)	Variant calling and annotation
BCF (.bcf)	Binary Call Format	Compressed binary version of VCF	Efficient variant data storage
GFF (.gff)	General Feature Format	Stores genome annotations	Gene structure and features
GTF (.gtf)	Gene Transfer Format	A more detailed version of GFF	Genome annotation and gene structure
BED (.bed)	Browser Extensible Data	Stores genomic regions (e.g., exons, promoters)	Defining genomic intervals
TSV (.tsv)	Tab-Separated Values	Stores structured tabular data	Gene expression and metadata
CSV (.csv)	Comma-Separated Values	Stores tabular data with comma separation	Experimental data storage
PDB (.pdb)	Protein Data Bank	Stores 3D protein structures	Protein structure analysis
MOL (.mol, .sdf)	Molecular Data Format	Stores chemical compounds and molecular data	Chemical informatics and drug discovery
CIGAR String	Compact Idiosyncratic Gapped Alignment Report	Describes how a read aligns to a reference genome	Used in SAM/BAM for alignment representation
MAF (.maf)	Mutation Annotation Format	Stores somatic mutations in cancer studies	Cancer genomics and mutation annotation

GA4GH Formats	Global Alliance for Genomics and Health Formats	Standardized formats for genomic data sharing	Genomics data interoperability
------------------	---	--	-----------------------------------