

# MPEG-G allows more than 99% faster initial processing compared to legacy formats

## How does the Selective Access work?

The Selective Access is based on MPEG-G's access unit structure and embedded indexing. One enables focusing on specific genetic targets without accessing and decoding the entire data. The other strongly accelerates data processing in genomic pipelines by completely removing the need for pre-processing stages and making parallel processing and data flow possible.

## What is MPEG-G?

MPEG-G (ISO/IEC 23092) is an ISO international standard for the representation of genome sequencing data and associated metadata.

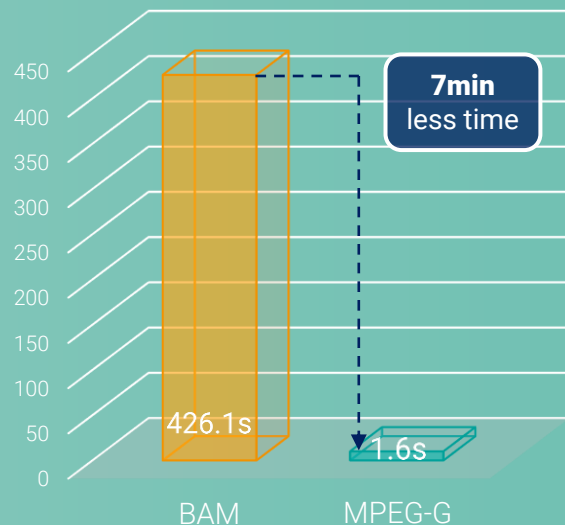
The standard MPEG-G aims to provide a framework for developing interoperable applications towards genuinely efficient and economical handling of genomic information.

## Use-Case: Selective Access Benefit

The total processing time – sorting, indexing, and accessing the data – for multiple regions of the *CFTR* gene amongst a Whole-Exome Sequencing (WES) dataset for a BAM format file needs 426,1 seconds - 7 minutes and 6 seconds. The MPEG-G formatted file was processed in 1.6 seconds, which amounts to 99.6% of less time spent using the MPEG-G format.

**GenomSys professional solution offers substantial time savings with more than 99% less time spent on processing MPEG-G formatted datasets.**

Comparison for one WES genomic dataset in BAM and MPEG-G.\*



\* The BAM file processed with samtools 1.11 using the commands 'samtools sort -@ 4 <bamfile> -o <sortedbamfile>', 'samtools index -@ 4 <sortedbamfile>', 'samtools view -@ 4 -L <bedfile> -M <sortedbamfile> -o output.sam' for 27 regions of the *CFTR* gene. Sorting time for MPEG-G included in access time, as sorting is performed concurrently with data access (even though not strictly required by MPEG-G to index and perform random access). Selective access time including sorting and indexing time.