

Benchmarking End-to-End, On-Premise, WGS Analysis, using Sentieon DNAscope Secondary Analysis with OmniTier Insight Tertiary Analysis

Introduction

As sequencing technologies become faster and more affordable, a shift toward whole-exome sequencing (WES) and whole-genome sequencing (WGS) is emerging. These approaches offer broader insights, enabling the discovery of novel genetic variants and a deeper understanding of complex traits, transforming research and personalized medicine in the genomic era.

Bioinformatics tools are also evolving to meet the requirements of challenging clinical applications. Initially it was necessary to analyze whole genome data in the cloud, to take advantage of the massive compute resources available. However, as on-premise compute processing and AI technology has advanced, end-to-end whole genome analysis is now available.

This application note details the benchmarking results of a whole genome data processing software pipeline comprising of Sentieon DNAscope, secondary analysis, followed by OmniTier Insight Hereditary, tertiary analysis, all running on OmniTier's CompStor Workgroup, on-premise server. These software tools are designed to meet the high-performance requirements of whole genomes, as well as whole exomes and panels.

About Sentieon DNAscope

DNAscope uniquely integrates haplotype-based variant calling with machine learning to enhance accuracy. As the successor to GATK HaplotypeCaller6, DNAscope retains a similar logical framework while introducing key improvements in active region detection and local assembly, significantly boosting sensitivity and robustness, particularly in high-complexity regions.

By leveraging a machine learning model, DNAscope generates candidate variants with additional informative annotations. These annotated candidates are then processed through the model for variant genotyping, leading to

Highlights

- Sentieon DNAscope secondary analysis pipeline provides accurate and efficient WGS germline mapping and variant-calling.
- OmniTier Insight Hereditary tertiary analysis provides fast, accurate germline, hereditary and rare disease variant analysis.
- Integration of both Sentieon DNAscope and OmniTier Insight onto a single, on-premise, CPU based server, is a fast, efficient solution for both long read, and long read WGS analysis.

greater accuracy in both variant calling and genotyping. DNAScope LR module further provides capability to process long reads as well.

With advancements in genomic data processing algorithms and a highly optimized implementation, DNAScope and DNAScope LR pipelines offer alignment, pre-processing, QC, variant calling and filtering functions, deliver results 5 to 10 times faster than the best opensource pipeline, offering both speed and precision in variant analysis.

About OmniTier Insight

There are 2 variants of Insight tertiary analysis:

- i) Insight Oncology, and
- ii) Insight Hereditary

Insight Oncology is tertiary analysis software tool that performs AI driven variant interpretation on either tumor or tumor+normal variant data, from gene panels to WES or WGS samples. Outputs include variant oncogenicity classification and reports containing AMP classification with associated drugs and therapies. For tumor+normal analysis, detection of hereditary cancer related variants is also supported.

Insight Hereditary is germline, hereditary and rare disease tertiary analysis software that runs on OmniTier's CompStor Server platform, either on-premise, in an HPC or in a private cloud. Insight Hereditary can analyze single, duo, trio or larger family collections of gene panels, WES or WGS. Outputs can either be draft or final causal reports of prioritized, classified, pathogenic and likely pathogenic variants. Alternatively, cohorts of samples can be analyzed using genome-wide associate study features.

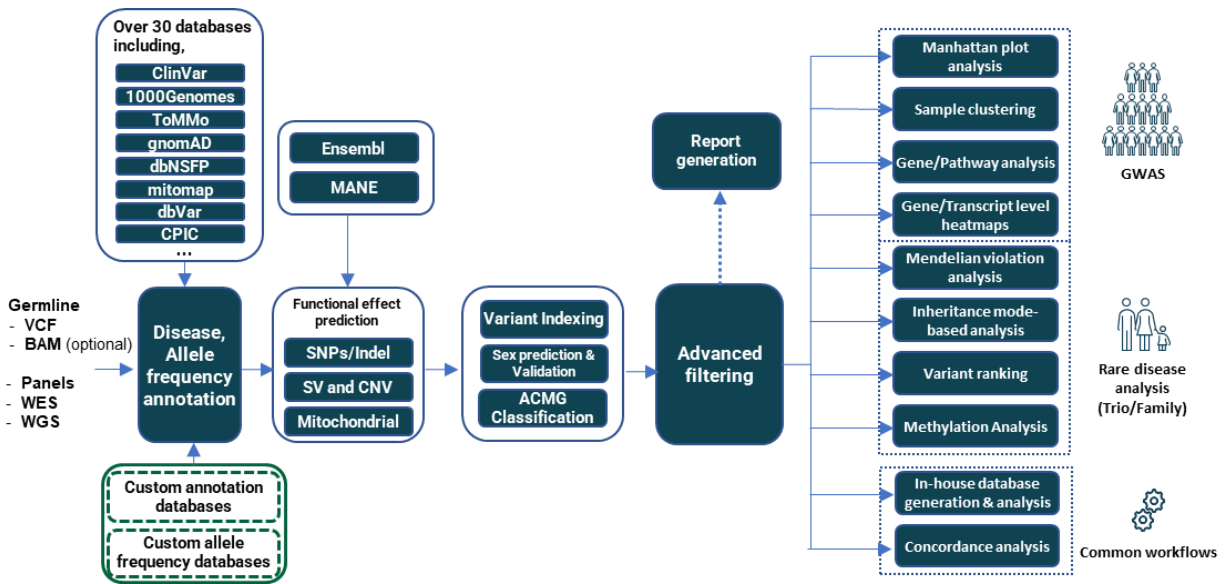


Figure 1: Insight Hereditary features

About OmniTier CompStor® Workgroup Server

The OmniTier high-performance 1U rack server was designed for demanding workloads in modern data centers. Powered by the latest Intel Xeon processors, it offers exceptional scalability, storage options, and advanced management features. Ideal for virtualization, database applications, and cloud services, it ensures reliability and efficiency for enterprise operations. The CompStor Workgroup Server is scalable in configurations of 1, 2, 4, or 8 nodes per compute cluster. Cluster nodes communicate with each other with a 10 GB/s network; the customer network can be connected at 1-10 GB/s speeds to transfer workflow input and output files and allow access to the control GUI.

Key server features

- 1U rack mount server
- Intel Xeon Gold 16 Cores, 3rd generation, 2 sockets

- 512 GB DRAM
- 12TB NVMe SSD
- 2x 10GBase-T network ports.

Methods

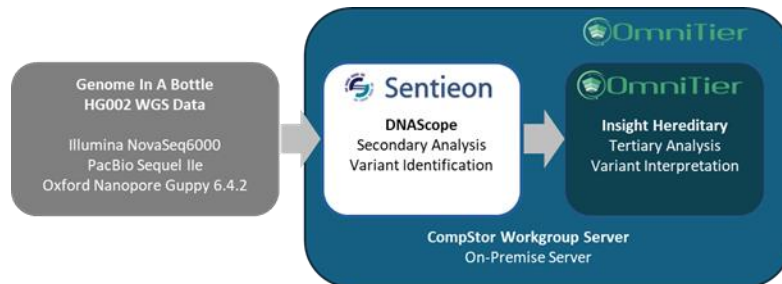


Figure 2. Setup of end-to-end WGS analysis benchmarking solution

Secondary Analysis using DNAScope

The correct version of DNAScope (version 202308.03) was downloaded as a binary from the Sentieon website and installed in a container on the OmniTier CompStor Workgroup server. A localhost trial license was issued to OmniTier for purposes of this benchmark testing. Sequencer specific scripts provided by Sentieon were then run for analysis purposes.

Tertiary Analysis using Insight

DNAScope generated VCF output files were then tested in Insight Hereditary as follows:

Importing VCF Files: VCF files generated by DNAScope or any other secondary analysis tool can be imported through YML/CSV files.

Analysis Workflow: After importing a VCF data file into CompStor Insight, analysis of samples can be performed including:

- Annotating raw variants with genetic traits and disorders, population allele frequencies, functional effect prediction and other useful information
- Filtering variants according to a comprehensive list of criteria
- Classifying variants according to ACMG guidelines
- Prioritizing variants based on significance

Loading Analysis Results: Once an analysis is completed, users can load the results by clicking on the icons in the "Action" column. This will display a table containing all the analyzed variants as shown in Figure 3.

Variant Details Page: To review the details of each variant, right-click on any variant row in the variant table to bring up the Single Variant Details Page. These details include allele frequency, functional effects, ACMG classification, a genome viewer (that automatically zooms to the selected variant) etc. An example of the single variant details page is shown in Figure 4.

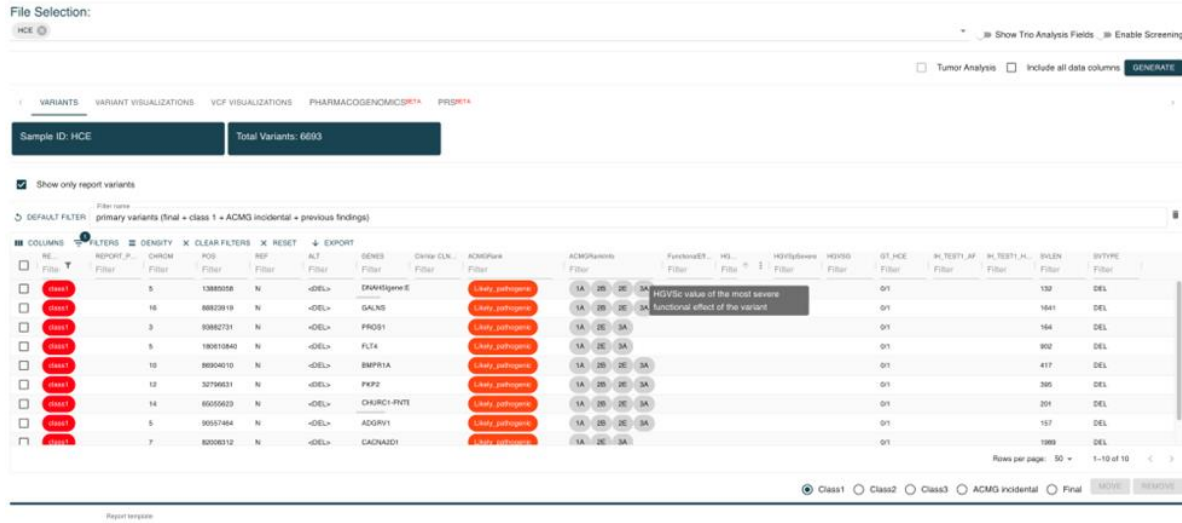


Figure 3: Review of analysis results through the variant details page

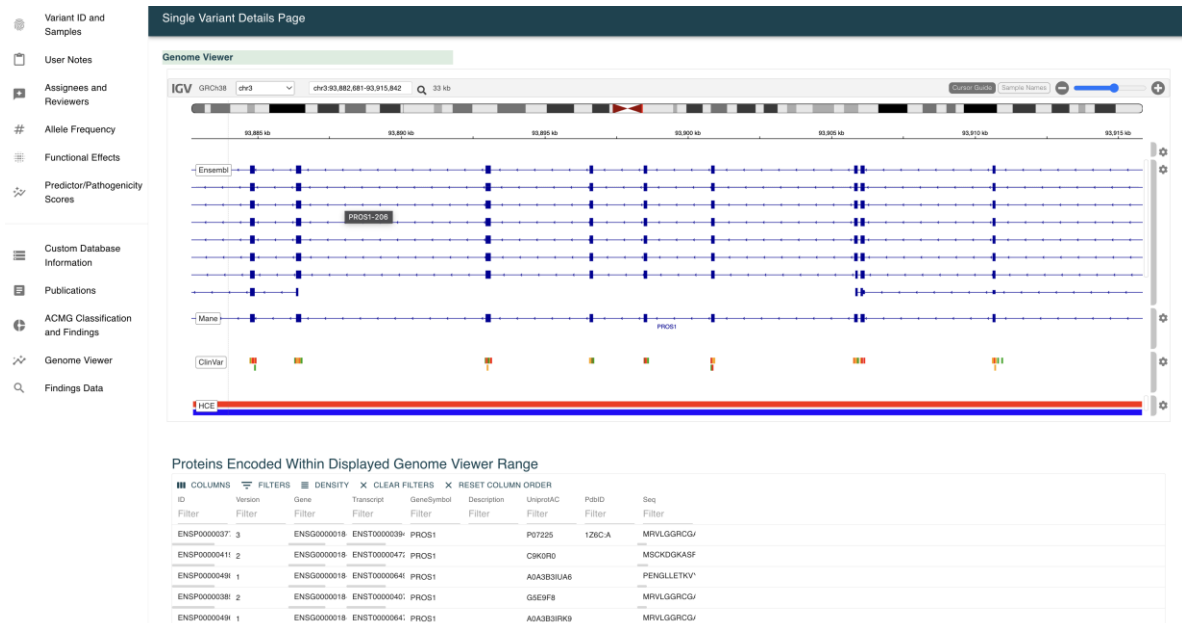


Figure 4: Single Variant Details Page along with integrated genome viewer.

Genome in a Bottle (GIAB) Test Data

The Genome in a Bottle Consortium is a public-private-academic consortium hosted by NIST to develop the technical infrastructure (reference standards, reference methods, and reference data) to enable translation of whole human genome sequencing to clinical practice and innovations in technologies. The priority of GIAB is authoritative characterization of human genomes for use in benchmarking, including analytical validation and technology development, optimization, and demonstration.

The GIAB HG002 genome is a high-quality human reference genome derived from a well-characterized individual from the Ashkenazi Jewish population, part of the Ashkenazi Trio (HG002, HG003, HG004). It serves as a gold standard for evaluating and benchmarking genome sequencing technologies, variant calling methods, and bioinformatics pipelines. HG002's genome has been extensively sequenced using

various technologies, such as Illumina, PacBio, and Oxford Nanopore, to achieve comprehensive coverage of single-nucleotide variants (SNVs), insertions and deletions (indels), and structural variants.

HG002 versions were used for benchmarking that had been sequenced with Illumina, PacBio, and Oxford Nanopore sequencers, to get a range of overall and independent runtimes, along with secondary analysis F1 variant calling quality scores.

Results

Runtime results were measured for the end-to-end pipeline, from FASTQ or unaligned BAM input to DNAscope and through Insight tertiary analysis, to produce a draft report containing a small list of filtered, prioritized, candidate variants. The results are as follows:

Dataset	Sequencer	Coverage	End-to-end (FASTQ to draft report) Runtime
HG002	Illumina NovaSeq 6000	35x (125.36 Gb)	115 minutes
HG002	PacBio, Sequel IIe	35x (108.62 Gb)	71 minutes
HG002	ONT, Guppy6.4.2	33x (103.23 Gb)	168 minutes

Of the total end-to-end runtimes quoted above, the DNAscope secondary analysis performance, from FASTQ to generation of variant VCF and BAM files, running on the CompStor Workgroup Server, performed as follows:

Dataset	Sequencer	Coverage	DNAscope Runtime	DNAscope F1 Score	Additional Info on the dataset
HG002	Illumina NovaSeq 6000	35x (125.36 Gb)	104 minutes	0.9961	PrecisionFDA Truth Challenge V2
HG002	PacBio, Sequel IIe	35x (108.62 Gb)	62 minutes	0.9987	PrecisionFDA Truth Challenge V2
HG002	ONT, Guppy6.4.2	33x (103.23 Gb)	160 minutes	0.9681	https://github.com/GenTechGp/gtqseq/blob/main/docs/data.md#na24385-hg002-promethion-data-30x

Of the end-to-end runtimes quoted above, the Insight tertiary analysis performance, from raw VCF file to generation of a filtered, prioritized list of candidate variants, running on the CompStor Workgroup Server performed as follows:

Dataset	Sequencer	Tertiary Analysis (vcf to draft report) Runtime
HG002	Illumina NovaSeq 6000	11 minutes, 33 seconds
HG002	PacBio, Sequel IIe	8 minutes 44 seconds
HG002	ONT, Guppy6.4.2	7 minutes 51 seconds

Summary

Integration of the Sentieon DNAscope secondary analysis software into OmniTier's CompStor Workgroup server enabled end-to-end analysis of Illumina, PacBio and Oxford Nanopore WGS data, on a single mid-range CPU based rack server.

The speed optimization during the development of both DNAscope and Insight results in short end-to-end runtimes, whilst keeping accuracy high. These good test results demonstrate that this is a good solution for those customers seeking an on-premise, end-to-end analysis solution.

OmniTier's Analysis Product Portfolio

CompStor Server Platforms



CompStor Server Pro



CompStor Server Workgroup



CompStor Server Personal

Essential Software

(Pro, Workgroup, Personal)

Novos *Secondary Analysis*

- Whole Exome, Whole Genome
- NGS and Long Read Sequencers
- SNP, indel, SV, CNV, Mitochondrial

Insight Hereditary *Tertiary Analysis*

- Hereditary and Rare Disease, GWAS
- Germline Panels, WES, WGS
- Single, Duo, Trio, Family

Insight Oncology *Tertiary Analysis*

- Tumor only, Tumor with matched normal samples
- Biomarker discovery

Contact OmniTier for more information.

Website: <http://www.omnitier.com>



Email: justin.cowling@omnitier.com

Contact Sentieon for more information.



Website: <https://www.sentieon.com>

Tel: 650-254-6360

Email: info@sentieon.com