OmniTier

COMPSTOR NOVOS®

High Throughput, Genome Secondary Analysis Appliance

Product Brief

For Research Use Only. Not for use in Diagnostic Procedures.

High-Throughput Genome Bioinformatics Appliance

Primary Analysis

Input: Raw detector signals

Output: FASTQ file with sequencing reads and quality scores

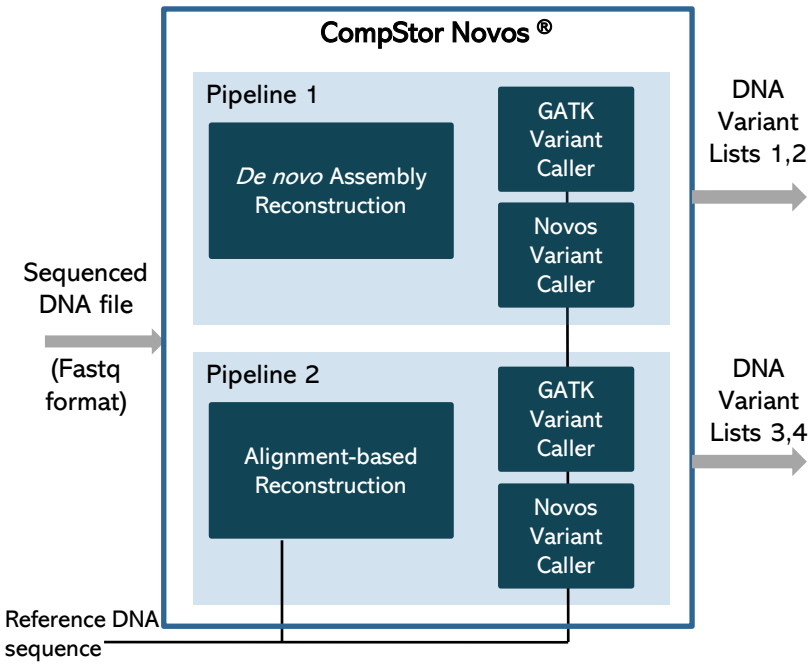
Secondary Analysis

Input: FASTQ files

Output:
i) BAM file with aligned reads
ii) File with Variant calls

Tertiary Analysis

Diagnosis and interpretation of variants and results



POWERING PRECISION MEDICINE

- ✓ **EXPANDED Variant Discovery**
- ✓ **ACCURATE Variant Calling**
- ✓ **HIGH-SPEED Run-Times**
- ✓ **AUTOMATED Pipelines**

Overview

CompStor Novos®. is a Whole Genome Sequencing (WGS) informatics pipeline solution with highest accuracy, lowest latency, and lowest cost per genome bioinformatics appliance with fast turnaround time. Variant calling accuracy exceeds *precisionFDA Challenge* winning results in all variant categories. It is a dual-pipeline enabling *de novo* assembly and reference alignment based methodologies.

CompStor Novos®. is a scalable, multi-noded server cluster solution that achieves super-computer-class performance in computation times using tiered-memory algorithms. The cluster configuration can adapt to meet a prescribed range of throughput requirements. An application based job scheduler automates batch processing and allows for a customized analysis workflow.

CompStor Novos®. variant calling is driven by a domain specific, deep learning algorithm that outperforms GATK Best Practices as well as alternative open source pipelines.

Product Highlights

Automated, Intuitive Pipeline

Fully integrated pipeline appliance that is quick and easy to set up and run, with a simple user interface.

From unpacking the box to NGS or long read sequencing in less than an hour.

Somatic

Germline

WGS and WES

Fast Run-Times

De novo assembly and variant calling in industry leading times, 1-3 hours depending on configuration

Scalable Increase WGS throughput for de novo processed personal genomes. Standard de novo assembled genomes can realize 3-5X reduction in cost per genome.

Accurate Variant Calling

Surpass standard open source tools in SNV/SNPs, short Indels and CNVs with alignment-based approaches; additionally, detect structural variants from *de novo*-assembly output contigs and unique indels

Bridge the gap between alignment-based and assembly-based genome analyses.

Dual Pipeline

De novo Assembly Reconstruction

Alignment-based Reconstruction

GATK & OmniTier Proprietary Variant Caller

OmniTier Proprietary Somatic Variant Caller

Illumina, BGI and PacBio sequencing

High Coverage Genomes

Sequence coverage tested up to 800x and file sizes up to several terabytes with 8-node appliance

Flexible Data Import

Automated job scheduler and data ingress web application supporting: FASTQ files from external client or FASTQ files residing on the CompStor Novos® appliance node

Accelerated Preprocessing

Demux / bcl2fastq

PacBio CCS processing

Preloaded and custom reference bundles

Extensible Platform

Ability to add new features and performance enhancements

Tertiary analysis

Analytics

Custom services

CompStor® Hardware Platform supports all functionality and only requires software updates and upgrades

Variant Calling Performance Highlights

Accuracy in Short Variant Calling

OnimTier's CompStor Novos® bioinformatics appliance shows greater variant calling accuracy across all seven NIST Genome in a Bottle (GIAB) datasets than GATK Best Practices pipeline as measured by F1 scores and total errors. Alignment and Assembly results for HG001 are shown in Figure 1. F1 scores are calculated from the fraction of true variants detected (recall) and the fraction of the variants called that are true (precision). In addition, CompStor Novos® F1 scores are higher than all winning entries from the most recent PrecisionFDA Truth Challenge III.

Variant calling utilizes a domain-optimized deep learning methodology to produce fewer false positives and more true positives.

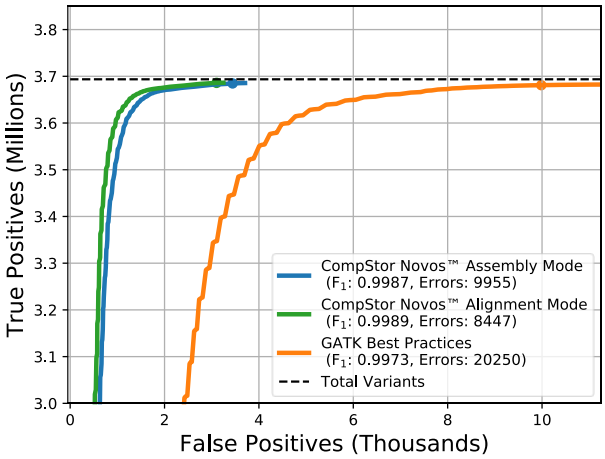


Figure 1 Receiver Operating Characteristics (ROC) for CompStor Novos® versus GATK Best Practices pipeline for HG001

Structural Variant Identification

The alignment-based approach is sufficient to call isolated and small variants. However, as variants become larger or clustered together, reads carrying their information become increasingly difficulty to place on the reference, resulting in missed variants in alignment-based pipelines. CompStor Novos®' Assembly overcomes this weakness because the genome reconstruction does not rely on a reference.

Contig length	Structural Variant	Chrom	Position	Left flank		Right flank		Affected gene
				Length	Mismatches	Length	Mismatches	
3086	17210 bp deletion	7	109453901	570	4	2509	7	---
2937	325 bp deletion	15	64633163	1031	1	1905	5	NG_051236.1
10974	542 bp insertion	8	27295979	5476	12	4955	15	NG_029510.1
4912	3630 bp insertion	4	97423310	659	4	628	2	---

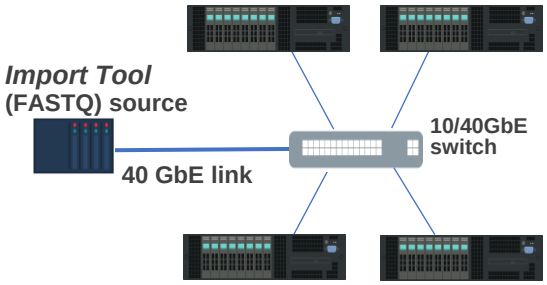
Table 1 Structural variants found with CompStor Novos® which were undetected with alignment methodology

Table 1 provides an illustrative example, where four structural variants undetected with alignment-based approach are found with the Novos pipeline. The left flank refers to the expected reference sequence on the left side of the variant and vice versa. Candidates for such variants are found by seeking reference matches in Novos' contigs at the SV flanks. Such SV candidates may then be confirmed by further manual analysis. Each of these SVs appears in the latest GIAB structural variant truth set. Where these variants are insertions, Novos' contigs exactly match the truth sets. Two of these variants affect the listed gene areas. For more details, please see OmniTier's academic paper at: https://www.biorxiv.org/content/10.1101/486092v1.

With its ability to enable variant detection over a wide range, CompStor Novo® offers an integrated platform for a wide variety of WGS applications.

CompStor Novos ® Appliance Configuration

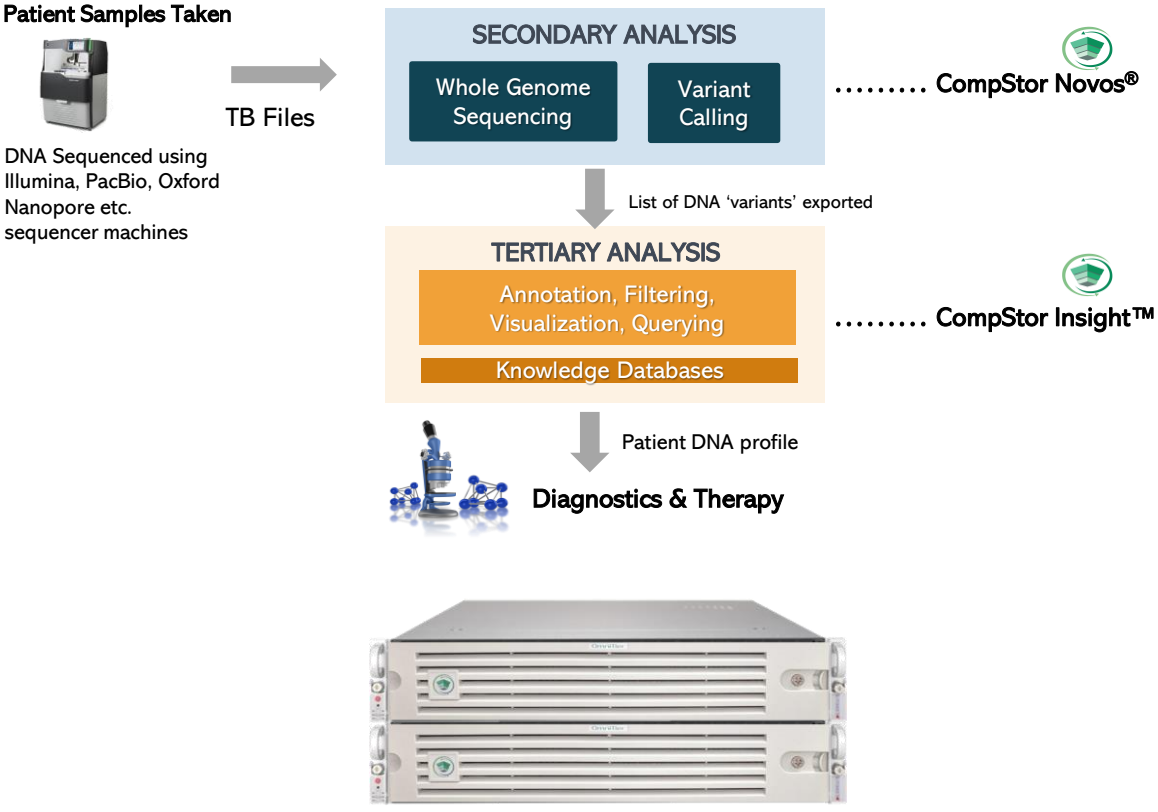
Options (1, 2, 3, 4+ nodes)



Assembly		Alignment	
Configuration	Run-time (hours)	Configuration	Run-time (hours)
2 nodes	3.0	2 nodes	**1.8**
4 nodes	2.0	4 nodes	**1.0**
8 nodes	1.0	8 nodes	**<1.0**

CompStor Novos® scalable, multi-node cluster — High speed data ingress, optimized memory tiers and multi-node communication drive *de novo* assembly and subsequent variant calling. Run-time estimates above are for 35x average coverage depth.

CompStor® Scientific Computing



For more information please email: sales@omnitier.com

For research use only. Not for use in clinical or diagnostic procedures.