OmniTier

COMPSTOR™ NOVOS

Whole Genome Sequencing Assembly & Variant Calling

Product Brief

High-Throughput Genome Bioinformatics

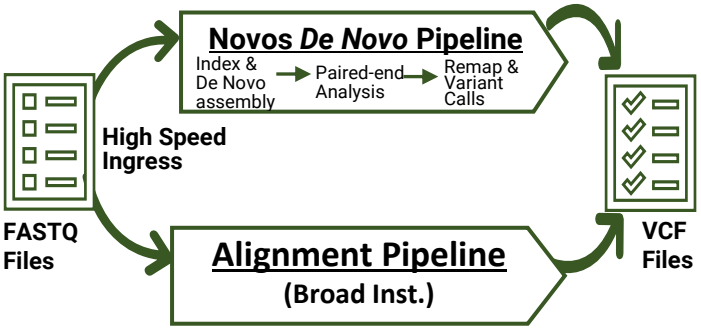
Overview

CompStor Novos is designed to meet the needs of high accuracy, fast turnaround time, and low cost in Precision Medicine. It is an integrated whole genome sequencing (WGS) bioinformatics solution that enables routine *de novo* assembly and associated variant calling. *De novo* assembly more accurately reconstructs the personal genome, enabling unique insight into complex variants, long insertions and deletions (Indels) and structural variations.

CompStor Novos is a scalable, multi-server cluster solution that achieves super-computer-class performance in *de novo* assembly computation times using tiered-memory algorithms. The cluster configuration can adapt to meet a prescribed range of throughput requirements.

While CompStor Novos enables the highly accurate assembly method in personal genome analysis, it also supports the standard alignment-based pipeline for variant calling, similarly optimized for multi-node operation.

Both pipelines are fully automated, taking short-read FASTQ input files from Illumina sequencers to output the variant call VCF files via an ease-of-use Import Tool.

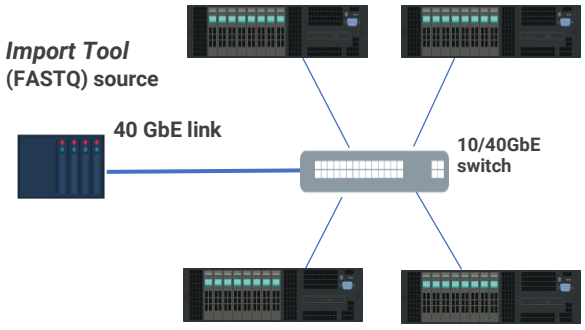


Analysis Pipeline Workflow – CompStor runs two separate pipelines: 1. Novos *de novo* assembly and variant call; and, 2. Alignment-based GATK Best Practices. A web application operates the cluster with automation and scheduling conveniences.

POWERING PRECISION MEDICINE

- ✓ ACCURATE Variant Calling
- ✓ HIGH-SPEED, SCALABLE Performance
- ✓ AUTOMATED, DUAL Pipeline support

Appliance Configuration



Configuration	Run-time (hours)
2 nodes	6
4 nodes	3
8 nodes	2

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 36x average coverage depth.

Product Highlights

Fast Run-Times

De novo assembly and variant calling in industry leading times, 2-6 hours depending on configuration

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

Dual Pipeline Support

De Novo & Alignment based (GATK Best Practices) pipelines supported on a single high throughput appliance

Automated, Intuitive Pipeline

One-button FASTQ → VCF automation makes it easy to use and increases workflow efficiency

Accurate Variant Calling

Match SNV/SNPs, and short Indels with GATK alignment-based approaches; additionally, detect clustered and structural variants from *de novo*-assembly output information

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

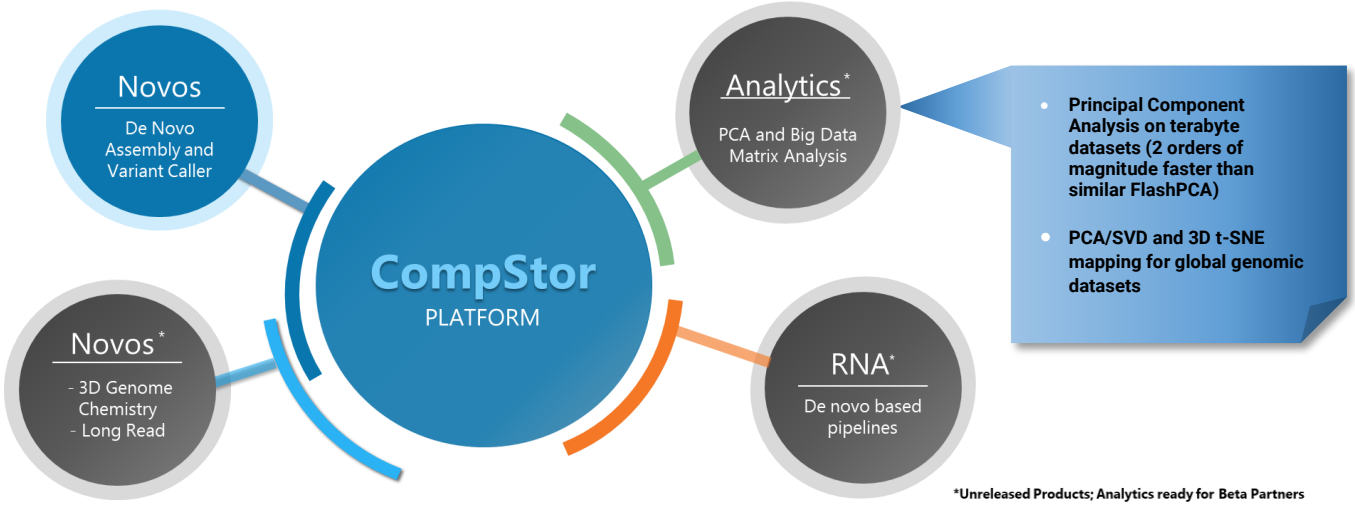
High Coverage Genomes

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 application supporting: FASTQ files from external client or FASTQ files residing on the CompStor Novos appliance node

The CompStor family extends to multiple applications in genomics and scientific computing



*Unreleased Products; Analytics ready for Beta Partners

Variant Calling Performance Highlights

Matching Alignment-based Variant Calling

Alignment-based methods rely on aligning reads to the reference genome for genome reconstruction and are widely used for variant calling. CompStor Novos performs the same function using de novo assembly, which aligns the same read data without a reference.

Four separate *Genome in a Bottle (GIAB)* datasets, with coverage depths ranging from 36x to 100x, were used to compare the truth sets from the two reconstruction methods using the same GATK variant caller from the Broad Institute. Figure 1 shows a representative comparison for the NA12878 genome dataset. More results are available from OmniTier's academic paper available at https://omnitier.com/resources/compstor. CompStor Novos compares well with the optimized alignment-based calling of point variants and short indels.

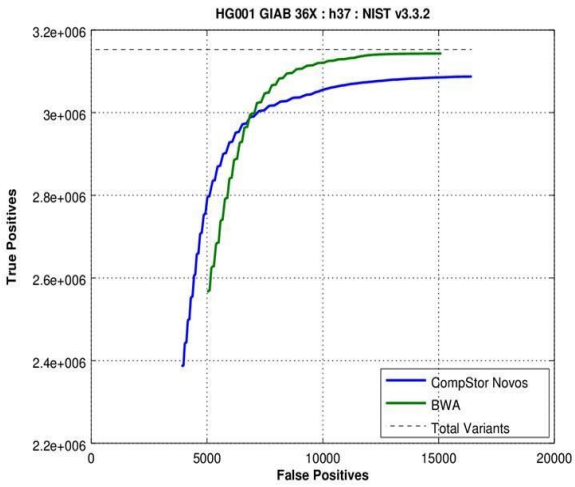


Figure 1 Receiver Operating Characteristics (ROC) for BWA method and CompStor™ Novos

Calling Structural Variants (SV)

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' approach is able to overcome 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 in BWA method

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://omnitier.com/resources/compstor.

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

Upgrade and Extendibility

CompStor Novos – May be upgraded with expansion of compute nodes for faster run-times and higher throughput.

	2-Node	4-Node	8-Node
Run-Time per Genome (@36x WGS)	6 hours	3 hours	2 hours
Daily Genome Throughput	4	8	12

CompStor Novos is extendible beyond variant calling to include further analysis on the VCF file. OmniTier's **CompStor Analytics** offers massively parallelized scientific computing solutions for machine learning algorithms including matrix analysis and linear regression. Datasets that are tens of terabytes can be analyzed in times comparable to super-computer-only solutions. OmniTier would collaborate with partners to integrate the analytics modules for their applications.

Sequenced Dataset Support

Sequencers	Illumina NGS
Short read length	75-300 base pairs (constant)
Paired-end read support	Yes
Max Coverage	800x
Min Coverage	10x
Reference dataset:	GR37

Appliance Technical Specifications

Form Factor	2U rackmount per node
Dimensions	Width 17.2" (437mm); Height 3.5" (89mm); Depth 30.7" (780mm)
Processor	Intel Xeon® Gold
Processor sockets	4
Memory	512 GB DDR4 2600; variable NVMe SSD
Power supplies	Titanium efficiency 750W AC power supply; 1100W DC power supply; Platinum efficiency 495W, 750W, 1100W AC power supply
Operating system	CentOS
Software	CompStor Novos 1.0; ClusterMaster 1.0; GATK Haplotype Caller

Ordering and Support

Please contact sales@omnitier.com or +1-408-508-6471 for further information or a quotation.

Product	Part Number
CompStor™ Novos Appliance	CS-2000 (2-node); CS-4000 (4-node)
CompStor™ Novos Server Expansion (2-nodes)	CS-2000UP
CompStor Analytics Upgrade (Software and NVMe SSD)	CSA-2000; CSA-4000

Support is provided by OmniTier and its server partner SuperMicro Inc. Server hardware issues are addressed with varying tiers of onsite support, ranging from 4 to 48 hours. OmniTier provides complete bioinformatics pipeline support from service engineers and bioinformatics specialists.

- **4 to 48 hours onsite response**
- **24x7x365 call line addressing hardware incidents**
- **OmniTier ticketed based support and dedicated appliance engineers**