OmniTier

COMPSTOR NOVOS®

Whole Genome Sequencing Variant Calling Appliance

Product Brief

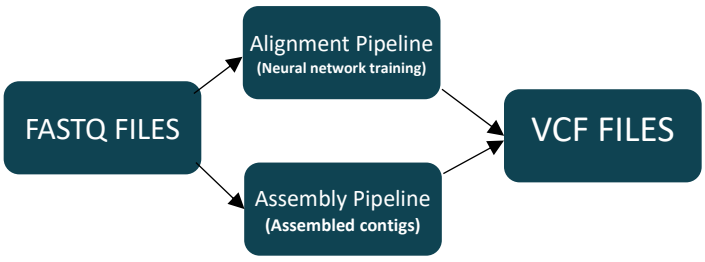
For Research Use Only. Not for use in Diagnostic Procedures.

High-Throughput Genome Bioinformatics

CompStor Novos® is a germline, whole genome sequencing (WGS) bioinformatics appliance designed for high accuracy, fast turnaround time, and low cost in Precision Medicine. 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 *de novo* assembly 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.

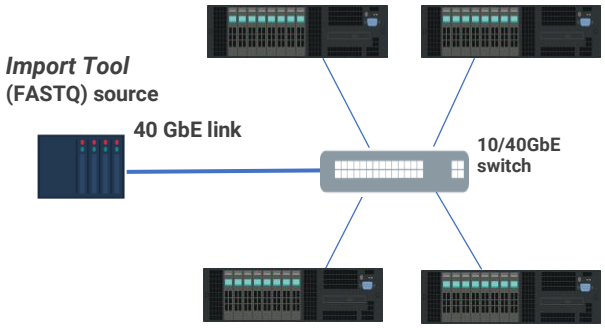


Analysis Pipeline Workflow — CompStor runs two separate pipelines: 1. Novos alignment and variant call; and, 2. Novos Assembly and variant call. A web application operates the cluster with automation and scheduling conveniences.

POWERING PRECISION MEDICINE

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

Appliance Configuration



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.

Product Highlights

Fast Run-Times

De novo assembly and variant calling in industry leading times, 1-3 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 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

Surpass standard open source tools in SNV/SNPs, and short Indels 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 with expanded variant

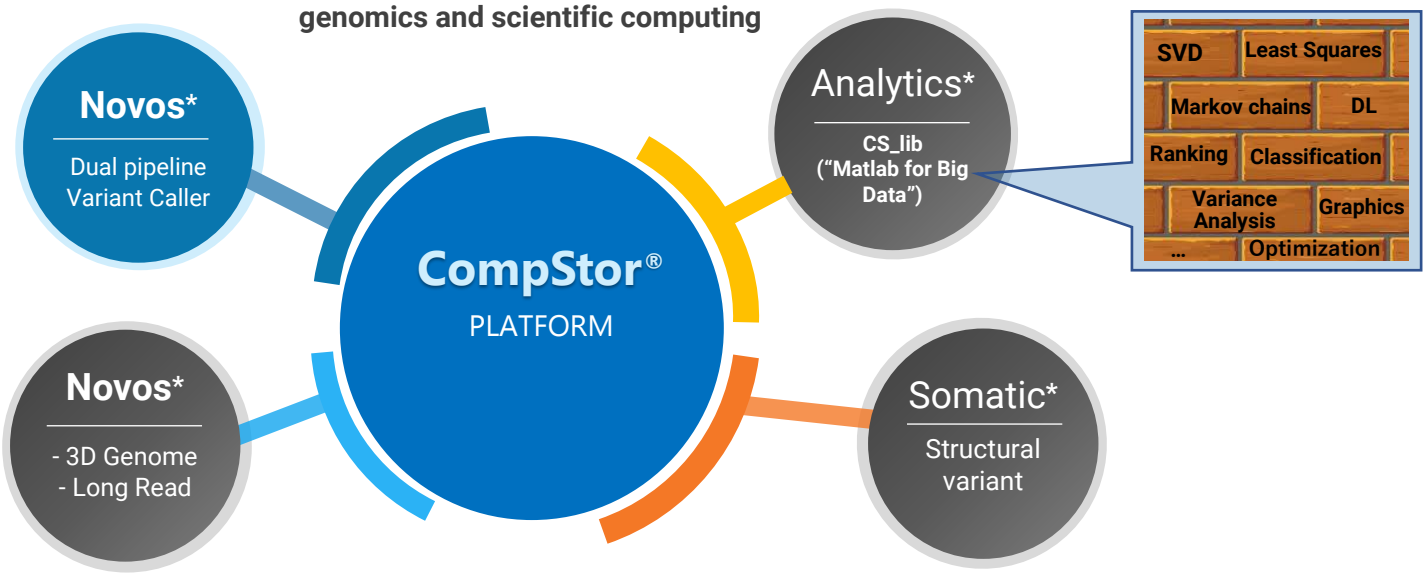
High Coverage Genomes

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

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



Variant Calling Performance Highlights

Accuracy in Short Variant Calling

OmniTier'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 F_1 scores and total errors. Alignment and Assembly results for HG001 are shown in Figure 1. F_1 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® F_1 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.

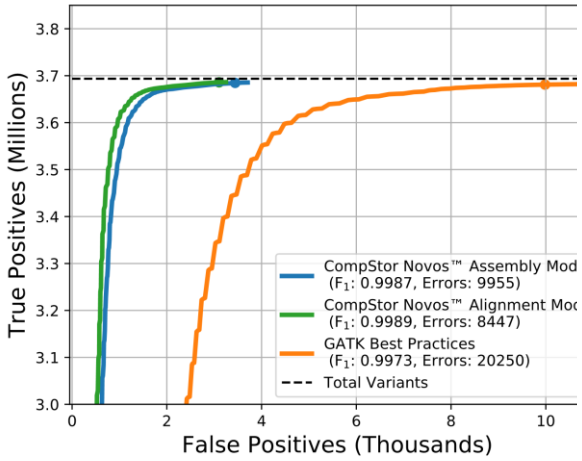


Figure 1 Receiver Operating Characteristics (ROC) for CompStor Novos versus GATK Best Practices pipeline for HG001

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 Novos™ offers an integrated platform for a wide variety of WGS applications.

Upgrade and Extendibility

CompStor Novos® – Add compute nodes for faster run-times and higher throughput.

Reconstruction Method (@35x)	2-Node	4-Node	8-Node
***De novo* Assembly (runtime)**	3 hours	2 hours	1 hour
Reference Alignment (runtime)	1.8 hours	1 hour	.66 hours

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; BGI/MGI
Short read length	75-250 base pairs (constant)
Paired-end read support	Yes
Max Coverage	800x
Min Coverage	10x
Reference dataset:	GR37 & GRH38

Appliance Technical Specifications

Form Factor	2U rackmount (All data per node)
Dimensions	Width 17.2" (437mm); Height 3.5" (89mm); Depth 30.7" (780mm)
Processor	Intel Xeon® Gold 6148 CPU @ 2.40GHz
Processor sockets	Quad (4); 80 Total Cores
Memory	768 GB DDR4; 12.8 TB SSDs (4x3.2TB)
Power supplies	Dual Titanium 1000W/1600W, Redundant Power Supplies
Power and operating conditions	AC input Voltage: 200-240V (Recommended) or 100-127V; Total Output Power: 1600W; Operating Temperature: 10C-35C (50-95F)
Operating system	Linux CentOS
Software	CompStor Novos® 1.0; ClusterMaster 1.0

Ordering and Support

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**
- **OmniTier ticketed based support and dedicated appliance engineers**

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

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