# BIOINFORMATICS APPLICATIONS NOTE # 2

*Competitive Review*

## Comparing OmniTier CompStor Novos® *Alignment* against top submissions in *precisionFDA Truth Challenge III*

OmniTier Inc. |  1591 McCarthy Blvd, Milpitas, CA 95035  |  2720 Superior Dr NW, Rochester, MN 55901

ABSTRACT

**Summary:** OmniTier's CompStor Novos® bioinformatics appliance, designed for Whole Genome Sequencing (WGS) applications, shows higher variant calling accuracy in recall and specificity than all five submissions of the SNP Recall and Indel Precision winner in the *precisionFDA Truth Challenge III*[1] as measured by $F_1$ score and total errors.

For HG001 CompStor Novos® produced 46-63% fewer errors than the five leading entries shown in Figure 1 and reported in the Challenge[1]. CompStor Novos® also runs in an accelerated configuration completing the analysis in 1 hour with a 2-node cluster.  Further run-time reduction is feasible with additional nodes.

## 1  INTRODUCTION

OmniTier's CompStor Novos® bioinformatics appliance uses a specialized memory architecture to drive fast run-times and accurate variant calling in germline, short-read WGS analysis. CompStor® tiered-memory architecture leverages optimized configurations of DRAM and NVMe SSDs for peak efficiency in the analysis of WGS data. The compute appliance supports an integrated dual pipeline, operable in either a de novo assembly or reference alignment mode via an automated web application based interface, outputting VCF files. This comparison utilizes Novos' *reference alignment* pipeline on a two-node cluster configured with 768 GB DDR4; 12.8 TB SSDs (4x3.2TB); and, 80 hyperthreaded cores per node.

Variant calling utilizes a WGS-optimized deep learning methodology to produce fewer false positives and more true positives and higher overall $F_1$ scores.

## 2  COMPARING DATA

Results from the precisionFDA Truth Challenge III were downloaded from the precisionFDA website[1]. The data was identical to the challenge conditions, 50X coverage datasets from HG001 and HG002 with NIST 3.2 truth sets.  The CompStor Novos™ algorithm was applied in a manner consistent with the precisionFDA Challenge rules.

| NIST Dataset | CompStor Novos® | Sentieon1 |
|:---:|:---:|:---:|
| HG001 | **.9994** | .9989 |
| HG002 | **.9993** | .9987 |
| $F_1$ Average | **.99935** | .9988 |
| Average Total Errors | **4,763** | 8,887 |

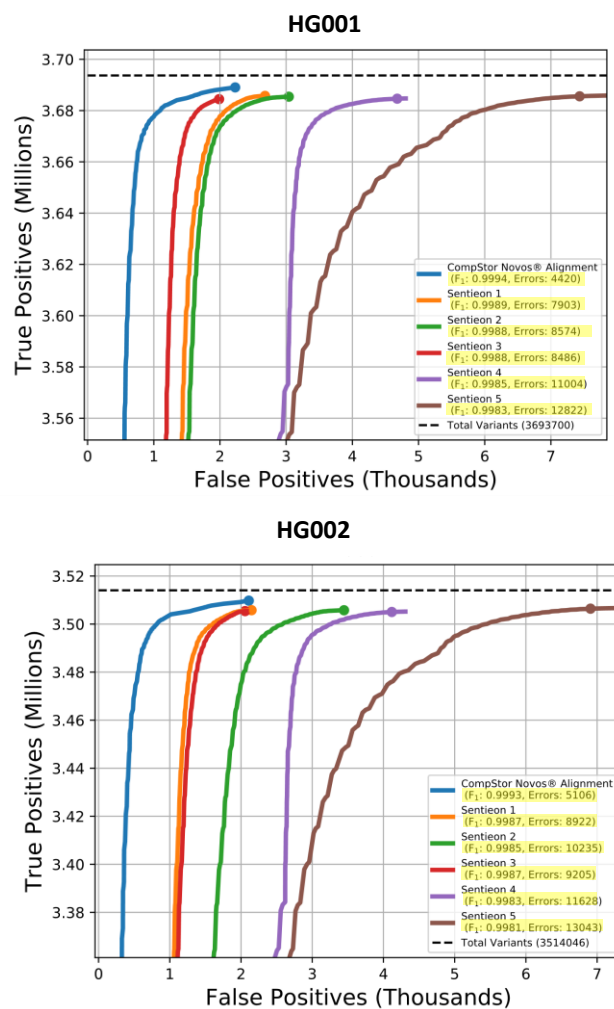**Table 1.** $F_1$ Scores for CompStor Novos® versus Sentieon[1] for HG001 and HG002



**Figure 1**. ROC Curves comparing true positives and false positives for CompStor Novos® Alignment and the 5 entries at PrecisionFDA Truth Challenge III for HG001 and HG002

REFERENCES
1 U.S. Food and Drug Administration. (2016) https://precision.fda.gov/challenges/truth/results. Retrieved from precisionFDA.