GENALICE MAP

WHY CHOOSE?
WHY CHOOSE?

No compromises. No choices. This was the challenging starting point when designing the new algorithms and turnkey configurations for our ultra-fast Next-Generation Sequencing (NGS) data processing and analysis software solution: GENALICE MAP.

MAP transforms NGS workflows into ultra-fast, simple and cost-effective processes. Our ground-breaking solution is also suitable for application in large scale, e.g. Whole Genome Sequencing (WGS), Transcriptome (WTS) or Exome (WES) projects.

MAP runs on commodity Intel Xeon E5 processors and is delivered on a turnkey (virtual) appliance, the GENALICE VAULT series.

GENALICE MAP has been thoroughly validated using various (benchmark) data sets in close collaboration with the world’s leading university medical centers and genomic organizations.

Described in this brochure are the performance data in head-to-head comparisons against some of the most widely-used alternative NGS analysis workflows.

The hardware configurations, input data and genome reference files used in these comparison studies were kept exactly the same for both workflows.
The performance of GENALICE MAP is driven by its proprietary newly designed algorithms, which are optimally geared towards modern hardware architecture.

As a result, MAP processes (map and call) NGS short reads up to over 150 times faster than widely-used conventional workflows. The exact speed gains in comparison to BWA-MEM and three different variant callers are displayed in the upper graph on the right.

On a simple Intel Xeon E5 2620 dual processor, the alignment of a complete human genome with 37x coverage takes around 25 minutes and variant calling takes just 5 minutes.
In head-to-head comparisons between GENALICE MAP and BWA-MEM in combination with three different variant callers; GATK HC, Platypus, and VarScan, on the platinum genome NA12878, it was shown that MAP outperforms the latter two on the \( F_1 \) scores. The \( F_1 \) score includes all variants (SNP and INDELs) and is a weighted average of sensitivity and precision.

An even more significant quality improvement can be seen on INDEL detection. GENALICE MAP can align reads with “infinite” gap sizes at high speed due to its architecture, algorithms, and comprehensive reference index. This enables it to detect longer INDELs with higher efficiency than other NGS workflows. MAP detects more long deletions than any of the other three NGS workflows.

<table>
<thead>
<tr>
<th>Workflow</th>
<th>Sensitivity (%)</th>
<th>Precision (%)</th>
<th>( F_1 ) Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENALICE MAP</td>
<td>99.26</td>
<td>99.29</td>
<td>0.99274</td>
</tr>
<tr>
<td>BWA-MEM/GATK</td>
<td>99.63</td>
<td>99.62</td>
<td>0.99683</td>
</tr>
<tr>
<td>BWA-MEM/Platypus</td>
<td>99.02</td>
<td>99.11</td>
<td>0.99063</td>
</tr>
<tr>
<td>BWA-MEM/VarScan</td>
<td>98.68</td>
<td>99.80</td>
<td>0.99235</td>
</tr>
</tbody>
</table>
SIZE MATTERS

GENALICE MAP allows researchers to generate a small footprint GAR (GENALICE Aligned Reads) file to replace BAM/SAM files produced in conventional workflows.

The size of a GAR file generated after alignment of a complete human genome 37x is only 5GB and thus offers a twenty-fold storage reduction. The file is ideal for ultra-fast variant calling afterwards and can be converted, if needed, to a conventional SAM/BAM file in near-time.

This significant file size reduction is achieved through a smart combination of reference backed encoding, base quality binning and best quality reference scoring.
HEAD TO HEAD COMPARISON

Using the popular benchmark Genome Comparison & Analytic Testing (GCAT) tool from bioplanet.com, we performed two head-to-head comparisons with widely used conventional workflows for NGS read alignment and variant calling. A full report is available at www.genalice.com.

The variant calls in these reports were been analyzed across a series of metrics including comparisons to the Genome in a Bottle (GIAB) call set, consistency with genotyping array data, and concordance with other variant callers.

GENALICE MAP detects SNPs and INDELs with a high sensitivity and precision rate, using the public benchmark high confidence SNP and INDEL call set (NIST v2.18) from the GIAB consortium. In both analyses, the exome datasets with 30 and 150 times coverage, MAP outperformed the different conventional workflows.

A significant difference can be observed in the 30x set, indicating a much higher sensitivity, while precision remains high, in areas with low coverage. In other words, GENALICE MAP not only offers a high speed and low cost NGS data processing solution, it also warrants improved outcome data.
EASE OF USE

GENALICE MAP has a short and simple workflow, consisting of only two processing steps: alignment and variant calling. Variant calling includes duplicate marking, INDEL realignment, genotyping, and filtering in a single processing step.

Conversely the BWA-MEM/GATK workflow is lengthy, complex and multi-staged; involving several software packages. The aligned reads are stored in the large storage footprint (~100GB) BAM format.

In addition, aligned reads must be prepared, which involves sorting, indexing, and duplicate marking, prior to variant calling. Each preparation step generates another BAM file. Moreover variant calling in GATK requires two independent processing stages: genotyping and filtering.

A SIMPLE TWO STEP WORKFLOW
GENALICE COMMITMENT

GENALICE is committed to improving the complete NGS data analysis workflow for as many different analysis jobs.

Our Production Box consists of two key components, which are required before any type of downstream analysis: mapping and calling.

The solution is built on general purpose hardware and also equipped with a GAR toolbox for conversion to SAM/BAM files, a GAR API, real-time monitoring software and 5TB of local storage.

This turnkey appliance, the GENALICE VAULT series, is available in different configurations for onsite and on cloud (AWS) usage. Modularly expansion of GENALICE MAP is in progress, to make it a complete NGS data Analysis Suite.

The following modules are already available or under development:
- Population Calling (PC)
- RNA-Seq data alignment and quantification
- Structural Variation analysis (SV)
- FASTA generation and injection (FASTA)
- Cohort Builder (CB)
- Somatic Calling, tumor/normal subtraction (SC)

A COMPLETE NGS DATA ANALYSIS SUITE
SIGNIFICANT SAVINGS

GENALICE MAP generates major gains in cost-effectiveness. Its substantial processing speed increase allows users to perform NGS data processing and analysis jobs on significantly smaller hardware configurations and in far less time.

MAP allows users to produce a small footprint GAR file to replace the BAM file. Consequently significant storage cost reductions can be achieved. In addition, this file size reduction allows organizations to replace the existing physical shipment of hard disks containing BAM files with secure network transfer.

This sums up to significant cost-savings, and reduced project risks. Lastly, better accuracy results in more cost-effective drug development, improved diagnosis and better treatment.
OUR COMPANY

GENALICE is a highly innovative biomedical big data solutions company, with global headquarters in the Netherlands.

We design and build groundbreaking software solutions for ultra-fast, highly accurate, and cost-effective DNA and RNA data processing and analysis.

The software is deployed on general-purpose hardware or flexible cloud. With GENALICE MAP, the company has introduced the first Next-Generation Sequencing (NGS) data processing and analysis solution with true population power.

By partnering with world-renowned research institutes and healthcare companies, GENALICE is committed to unlocking the potential of whole genome, exome and transcriptome sequencing for biomarker discovery and clinical application.

We design, build, sell and deliver all of our software solutions ourselves with a dedicated and highly motivated team of professionals.
A journey like ours can never be successfully done alone. We therefore carefully select product development, scientific, strategic and distribution partners from the ecosystem that we are active in.

A current status overview:
SUMMARY

In the overview below the key GAR file related features and overall user benefits of GENALICE MAP are depicted.