

# Helicube - Simple and Fast NGS Analysis

Helicube v1 is an integrated solution for managing and analyzing Next Generation Sequencing (NGS) data. Preinstalled with state-of-the-art software BALSAs and database.bio, and common pipelines such as BWA+GATK and TopHat, Helicube supports heterogeneous workloads, ranging from data analysis to results visualization. Helicube can serve multiple users via web browser. Helicube is designed to be HIPAA compliant; it allows users to share projects and data using carefully designed privileges, while strictly protecting private data. Helicube can scale up to multiple units to support the workloads of larger institutes; workload scheduling and balancing is also automatic, giving users convenience and efficiency.

## Highlights

**Fast and accurate genome analysis:** From raw reads to variants, BALSAs takes a few hours for WGS, and ten plus minutes for WES.

**Bioinformatics best practices:** Pre-installed pipelines include BALSAs, BWA+GATK, TOPHAT, STAR and Cufflinks. Extendable with Helicube Python SDK.

**Interactive variant analysis:** Variant interpretation and pathogenicity classification with over 30 databases including ClinVar, TCGA, ICGC, EVS and dbSNP.

**Easy sharing:** Fine-grained privileges enabling contributors, analyzers, methodologists, etc. to collaborate on your project

**Helicube User Guide:** [www.L3-Bioinfo.com/Products/Helicube](http://www.L3-Bioinfo.com/Products/Helicube)

**Demo account:** demo  
**Password:** KZ8a6LFGu  
Preloaded with projects, data and pipelines.

## Overview

Helicube v1 features two independent systems: the Helicube Engine for genome analysis and database.bio for visualization. The two systems can be used separately, or as an integrated unit.

The Helicube Engine stores and manages user data and analysis results securely. It enables management of large genomics projects through fine-grained access settings for collaborators. Top-notch tools and best practices for NGS analysis including BALSAs, BWA+GATK, TOPHAT, STAR and Cufflinks are pre-installed. As an advanced feature, Helicube provides an SDK for users to upload their own tools and create new pipelines.

database.bio is a clinical decision-support system that interprets genome variants (input with VCF and 23andMe formats). database.bio integrates genomic, clinical and pharmaceutical information from different sources into a uniform database in which variants are described with standard nomenclatures.

## Features

### BALSAs

BALSAs is an ultra-fast and accurate software for genome secondary analysis. Accelerated by a GPU, it is 5 or more times faster than BWA+GATK and can finish a WES analysis in minutes. It enables the application of NGS in the clinical context, where fast turn around time is required.

	BALSAs	BWA+GATK
WGS 38x (150 bp)	6 hours	> 30 hours
WGS 50x (100bp)	5 hours	> 25 hours
WES 210x (100bp)	12 minutes	210 minutes

Figure 1: Speed comparison of BALSAs and BWA+GATK on Helicube.

Pipeline	True+	False+	True-	False-
BALSA GCAT Illumina 150x	22,913	692	46,467,845	845
Bwamem+Gatk_HC_v3pt1	22,616	566	46,467,971	1,196
Bowtie2+Freebayes_Q40	22,567	1,474	46,467,063	1,158
isaacV01_13_06_20	22,533	743	46,467,794	1,254

Figure 2: Genome in a Bottle ([www.bioplanet.com/gcat](http://www.bioplanet.com/gcat)) benchmark (data from [www.bioplanet.com/gcat/reports/5236-lyjfvatdd/variant-calls/illumina-100bp-pe-exome-150x/balsa-gcat-illumina-150x/compare-571-2319-530/on-target/group-read-depth](http://www.bioplanet.com/gcat/reports/5236-lyjfvatdd/variant-calls/illumina-100bp-pe-exome-150x/balsa-gcat-illumina-150x/compare-571-2319-530/on-target/group-read-depth)).

Benchmarking on the quality of variants using NIST’s “Genome in a Bottle” demonstrates that the variants called by BALSA has the highest combined sensitivity and precision (refer to Figure 2).

### database.bio

database.bio delivers a broad set of features that allow users to quickly understand and visualize genome analysis results.

### Scalable

Helicube can scale up to more than ten hardware units to support the extensive workload of your institute. These units work

coherently in a management free and efficient manner, enabling you to focus on your research instead of the infrastructure.

### Managed and Monitored

Helicube only runs authorized tools, ensuring the integrity of the data handled by the system. It also provides a monitoring function to give users timely updates about system health.

### Data Protection

Helicube is designed to be HIPAA compliant and uses encryption standard AES256 for secure data transfer.

## About this Trial Helicube

This Trial Helicube comes with a single Helicube unit with ordinary configuration: 6 CPU-cores, 64GB RAM and 16TB hard disk storage (note that a Production Helicube will be equipped with at least 12 CPU-cores, 96GB RAM and also connected with a Network Attached Storage (NAS) or a user-provided storage system). As this Trial Helicube is without a NAS, it is recommended that users delete test data and results regularly and when prompted (available hard disk storage drops below 30%) to accommodate further testing. To keep the system robust, data upload or analysis scheduling will be disabled when available storage drops below 10%.

This Trial Helicube comes with a trial period of 21 days and requires Internet connection (note that a Production Helicube doesn’t require Internet connection). The trial is provided to demonstrate the capabilities of Helicube and to help in its assessment and evaluation. This Trial Helicube (both hardware and software) belongs to L3 Bioinformatics Limited and should be returned to L3 Bioinformatics Limited intact at the end of the trial period.

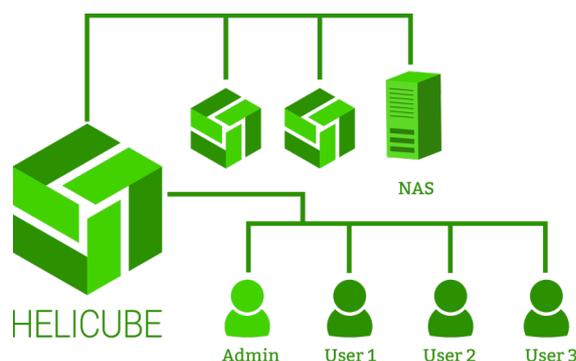


Figure 3: Helicube.