
OmicsBox
Bioinformatics Made Easy
biobam
BIOINFORMATICS SOLUTIONS

# OmicsBox Overview

Thousands of researchers trust OmicsBox for their daily bioinformatics tasks

OmicsBox is a leading bioinformatics software platform designed for end-to-end analysis of genomes, transcriptomes, and metagenomes. This user-friendly desktop application is efficient and powerful, enabling researchers to uncover valuable biological insights from complex omics data.

Trusted by top private and public research institutions worldwide, OmicsBox simplifies the processing of large datasets, streamlining the analysis workflow.

Its modular structure provides tailored tools and functions for diverse analyses, including de novo genome assembly, genetic variation analysis, differential expression analysis, and taxonomic classification of microbiome data.

OmicsBox also supports functional interpretation and offers rich data visualization capabilities.

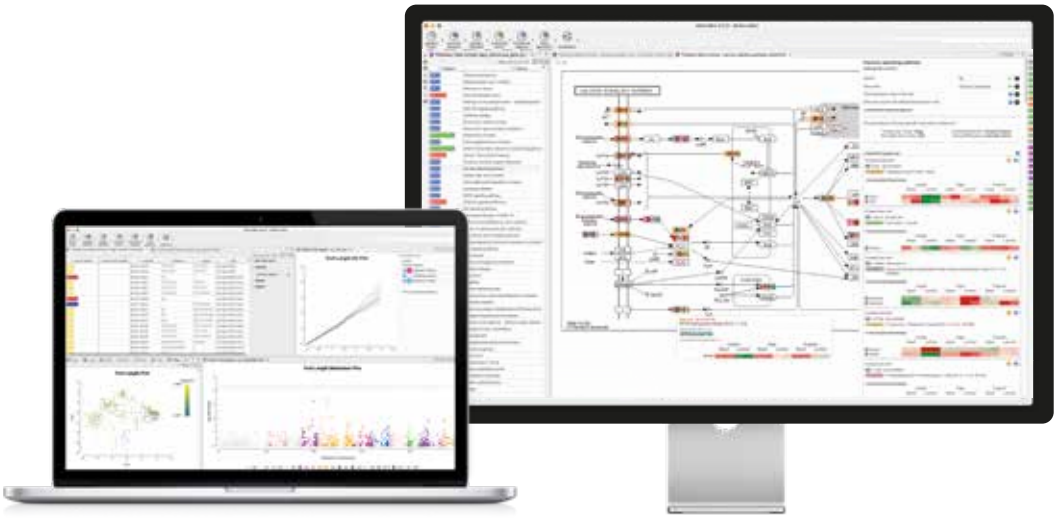
The functional analysis module, featuring the renowned Blast2GO annotation methodology, is particularly well-suited for non-model organism research — a fact supported by over 15,000 scientific citations.

OmicsBox is compatible with standard PCs and laptops running Windows, Linux, or Mac, offering a seamless experience right out of the box.

**Efficient**

**Powerful**

**Versatile**



# OmicsBox Modules

Five modules to easily process large and complex data sets

## Genome Analysis

- Quality Control and Assessment
- Genome Alignment and Assembly
- Repeat Masking and Gene Finding
- Genome Assessment and Curation Tools

## Transcriptomics

- De-Novo Assembly and Alignment
- Differential Expression Analysis
- Single-Cell RNA-Seq
- Long-Read Transcriptomics

## Metagenomics

- Metagenomic Assembly
- Taxonomic Classification
- Functional Annotation and Analysis
- Comparative Analysis

## Genetic Variation

- Fast Variant Calling and Filtering
- Supports GBS and WGS Data
- Variant Annotation
- Genome-Wide Association Studies

## Functional Analysis

- Comprehensive Sequence Analysis
- High-Throughput Functional Annotation
- Advanced Enrichment Analysis
- Combined Pathway Analysis

## General Features

- Manage Projects and Files
- Access to Cloud Computing
- Design, Run and Save Workflows
- Visualize Data, Genome Browser

### Intuitive Interface

OmicsBox features a user-friendly interface that simplifies navigation through datasets and tools, making complex data accessible to researchers of all backgrounds.

### Versatility in Analysis

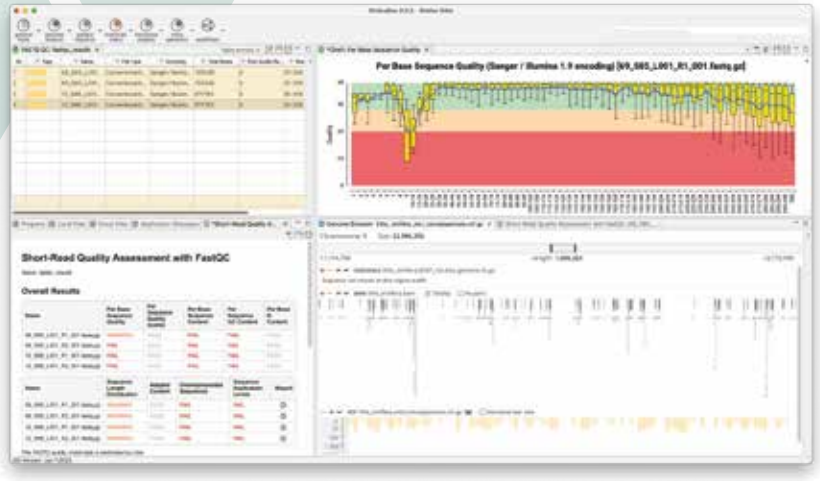
From alignments and functional annotation to differential expression analysis and pathway enrichment, OmicsBox encompasses a comprehensive range of analysis techniques. Having all these tools in one platform saves time and effort.

### Customization

Tailor your analysis to your specific needs with OmicsBox's flexible parameters and customizable workflows. These features enable you to refine your analysis, achieve accurate results, and handle even challenging datasets on standard equipment.

# Genome Analysis

## Convert raw DNA-Seq reads into a structurally annotated and curated draft genome



- Quality Control And Assessment
- De-Novo Assembly
- Long Reads Genome Analysis
- Alignment and Polishing
- Repeat Masking and Gene Finding

## Complete Genome Analysis for Short and Long Reads

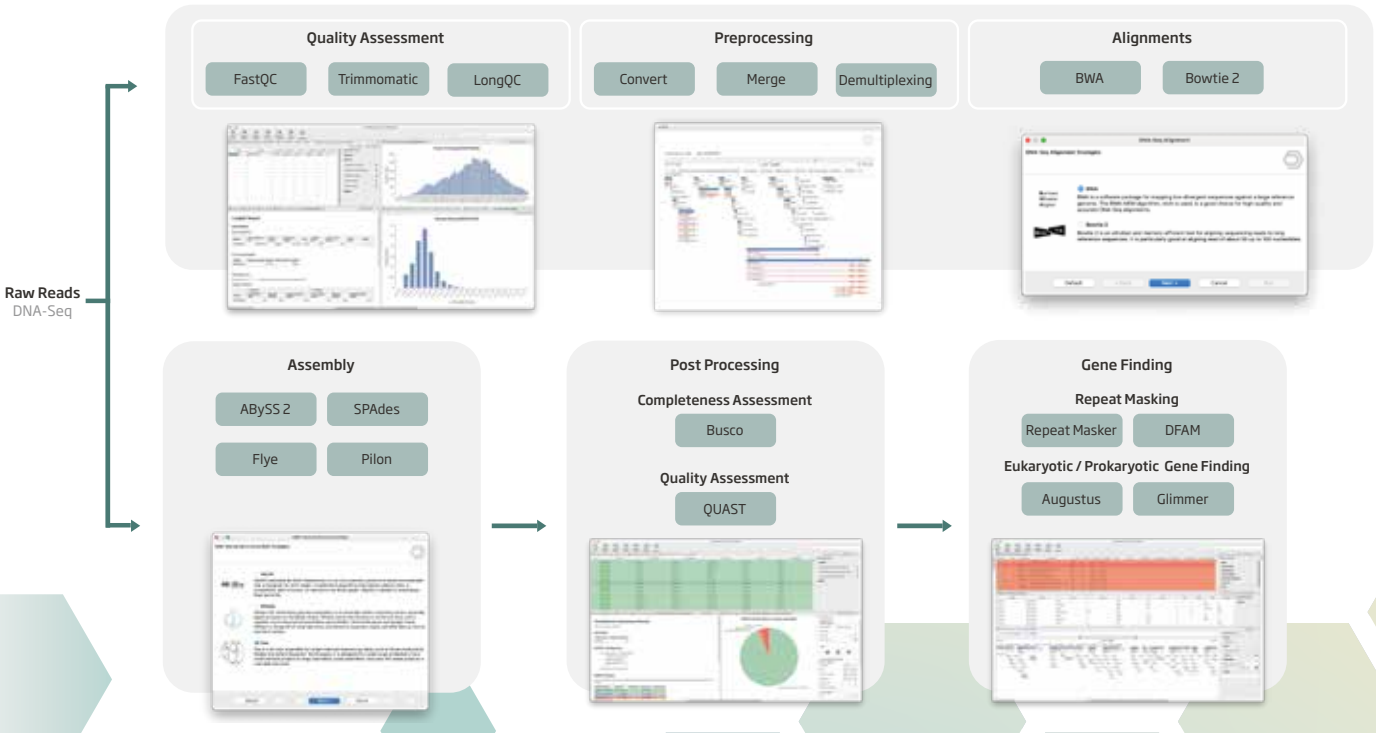
The Genome Analysis module facilitates the efficient characterization and analysis of newly sequenced genomes, guiding users seamlessly from raw reads to gene structures in a user-friendly interface.

Data quality control can be performed using tools like FastQC and Trimmomatic, enabling the filtering of reads and the removal of low-quality bases. For genome assembly, the module supports the reconstruction of whole genome sequences without requiring a reference genome or specialized hardware. It accommodates both short- and long-read sequencing technologies, utilizing popular algorithms such as **ABySS, SPAdes, and Flye.** Additionally, Flye can be paired with Pilon for sequence polishing to enhance assembly quality.

To ensure the completeness of the assembled genome, **BUSCO** is available for comprehensive assessment. Quast further allows for quality comparisons between multiple de novo assemblies, providing valuable insights into assembly performance.

For read alignment, the module offers efficient integration with **BWA and Bowtie2**, alongside tools for repeat identification and masking prior to gene prediction.

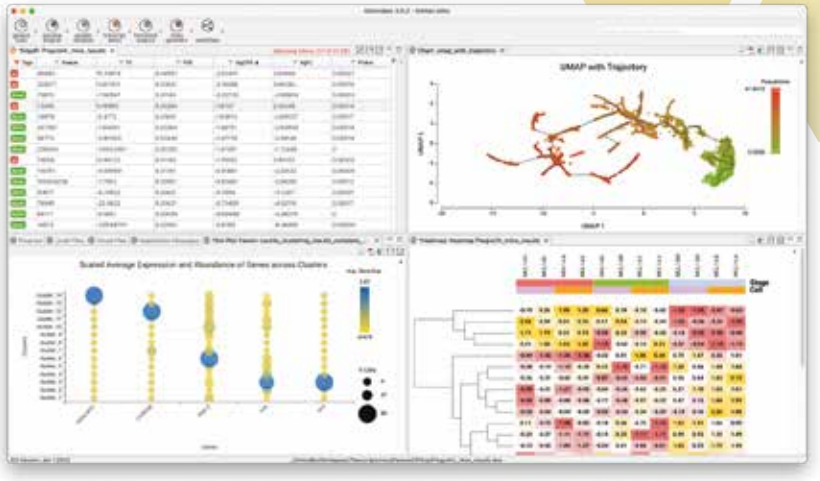
Finally, OmicsBox enhances genome visualization by enabling the integration of annotations as interactive tracks. These tracks combine genome sequences (.fasta), alignments (.bam), intron-exon structures (.gff), and variant data (.vcf), providing a comprehensive view of your genomic data.



# Transcriptomics

## Convert your RNA-Seq Samples into Biological Insights

- De-Novo Assembly
- RNA-Seq Alignment
- Differential Expression Analysis
- Long-Read Transcriptomics
- Single-Cell RNA-Seq



## Quality Control, Assembly, Quantification, and Differential Expression

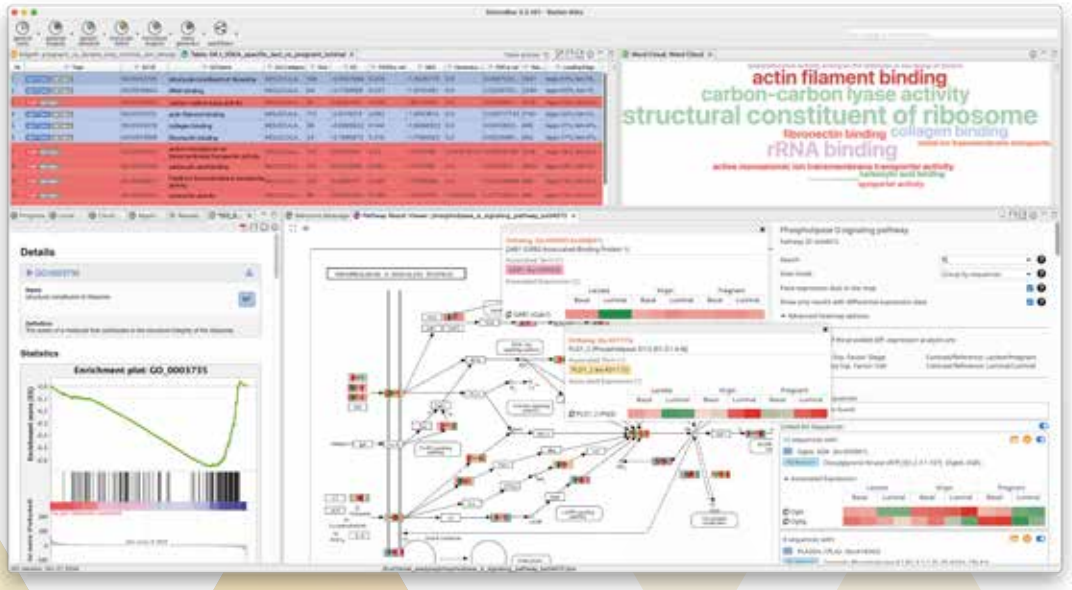
The Transcriptomics module provides a flexible and intuitive solution for processing RNA-Seq data, covering the workflow from raw reads to functional analysis. After quality control using FastQC and Trimmomatic, RNA-Seq data can be aligned to a reference genome with **STAR or BWA** or assembled de novo without a reference genome using **Trinity** to create a transcriptome. Additional tools for coding region prediction and coding potential assessment further refine the datasets.

The module supports expression quantification at the gene or transcript level, either with or without a reference genome, using tools like **HTSeq or RSEM.**

Comprehensive statistical charts offer insights into assembly, quantification processes, and result quality- Differentially expressed genes can be identified across experimental conditions or time points using statistical packages such as **NOISeq, edgeR, or maSigPro.** Results are presented in interactive heatmaps for intuitive interpretation.

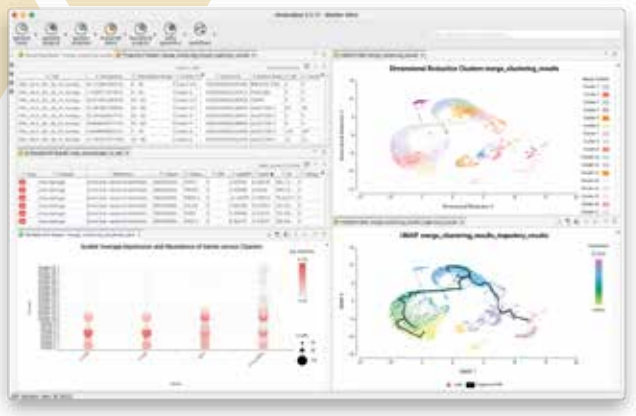
For **Long-Read** RNA-Seq data, transcriptomes can be reconstructed with various tools, both reference-based and de novo, and further characterized and filtered using SQANTI3.

**Single-Cell** RNA-Seq tools included in the module enable researchers to identify cell clusters, annotate cell types, and explore cell lineage trajectories in pseudotime. Differential expression and functional enrichment analyses are seamlessly integrated, making it easy to identify overrepresented and underrepresented biological functions. Additionally, the interactive UMAP/t-SNE plots facilitate straightforward data exploration and visualization.



# Single-Cell

## Exploratory Single-Cell Data Analysis Made Easy



OmicsBox simplifies single-cell RNA-Seq analysis by providing a streamlined and user-friendly solution, covering everything from raw scRNA-Seq reads to functional data interpretation.

With a strong focus on dynamic data exploration and interactive visualizations, it empowers researchers to efficiently analyze, interpret, and gain insights from even the most complex datasets.

## Features

**Quantification**

Leverage **STARsolo** to quantify scRNA-Seq reads. Compatible with both UMI-based technologies (e.g., 10x Chromium, Drop-seq) and full-length technologies (e.g., SMART-Seq, SmarTer), STARsolo allows you to customize quantification parameters for your specific dataset.

**Clustering**

Perform clustering with **Seurat,** including data preprocessing, correction, and integration. Analyze data from various samples, batches, or conditions to reveal meaningful biological patterns.

**Visualization**

Explore your data through rich visualizations such as violin plots, heatmaps, and bubble plots. Interactive UMAP and t-SNE projections enable dynamic exploration, allowing you to color cells by condition or gene expression and create personalized groups for detailed analysis.

**Cell Type Identification**

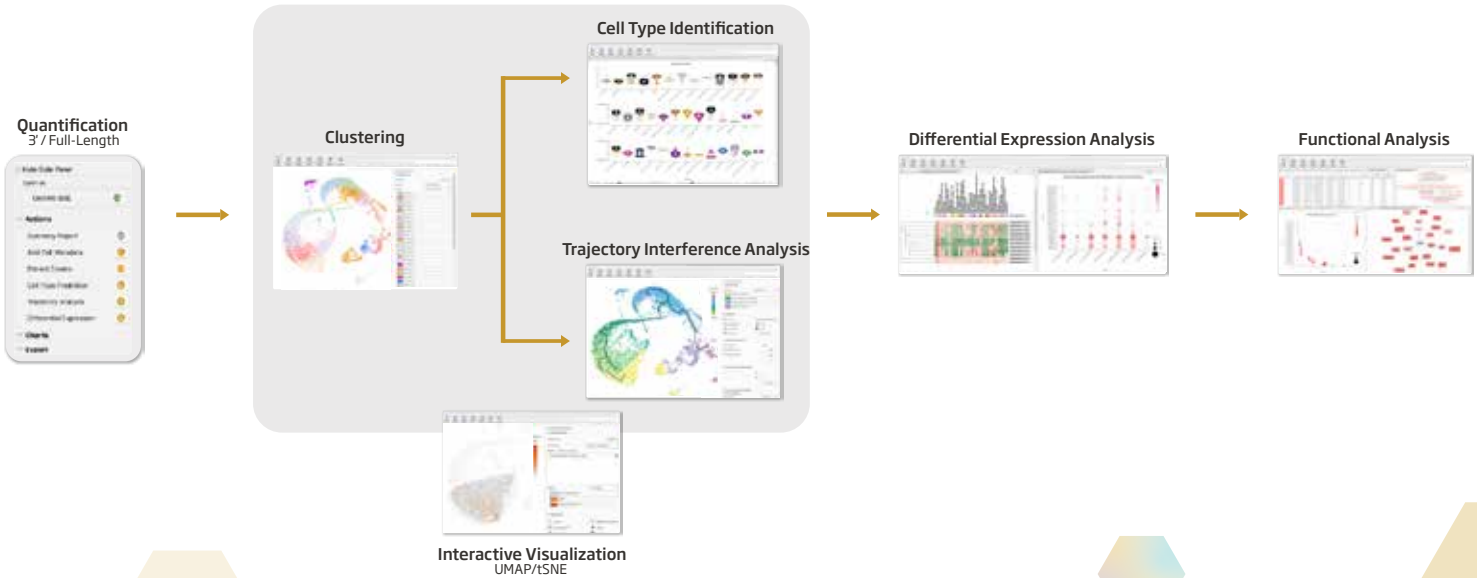
Identify cell types using **SingleR**, which compares your dataset with an annotated reference, or leverage **CellKb** for knowledge-based cell type predictions of the groups in your dataset. Evaluate the quality of predictions through detailed visualizations and gain confidence in your results.

**Trajectory**

Order cells along pseudotime with **Monocle3** to study differentiation processes and developmental stages. Conduct autocorrelation analysis to pinpoint the genes driving these trajectories, uncovering new biological insights.

**Differential Expression**

Compare gene expression levels across clusters, cell types, conditions, and pseudotime trajectories. Perform functional enrichment analyses to identify overrepresented pathways and gain deeper insights into the biological significance of your data.

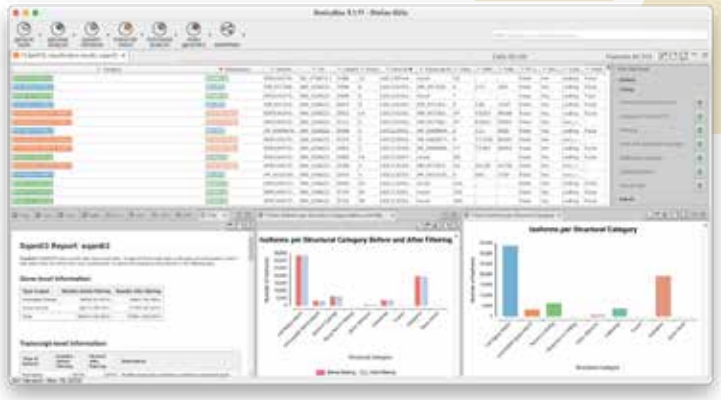


# Long-Reads

## Characterize and Curate Long-Read Transcriptomes with OmicsBox

Long-read technologies, such as PacBio and Oxford Nanopore, are ideal for characterizing and analyzing isoforms, as each read can represent an entire transcript.

With OmicsBox's suite of long-read tools, you can accurately reconstruct transcriptomes, assess their quality, and quantify isoforms with ease.



## Features

**minimap2**
Align your long reads using minimap2. While minimap2 offers numerous parameters, OmicsBox simplifies the process with predefined presets that can be tailored to your dataset for precise alignments.

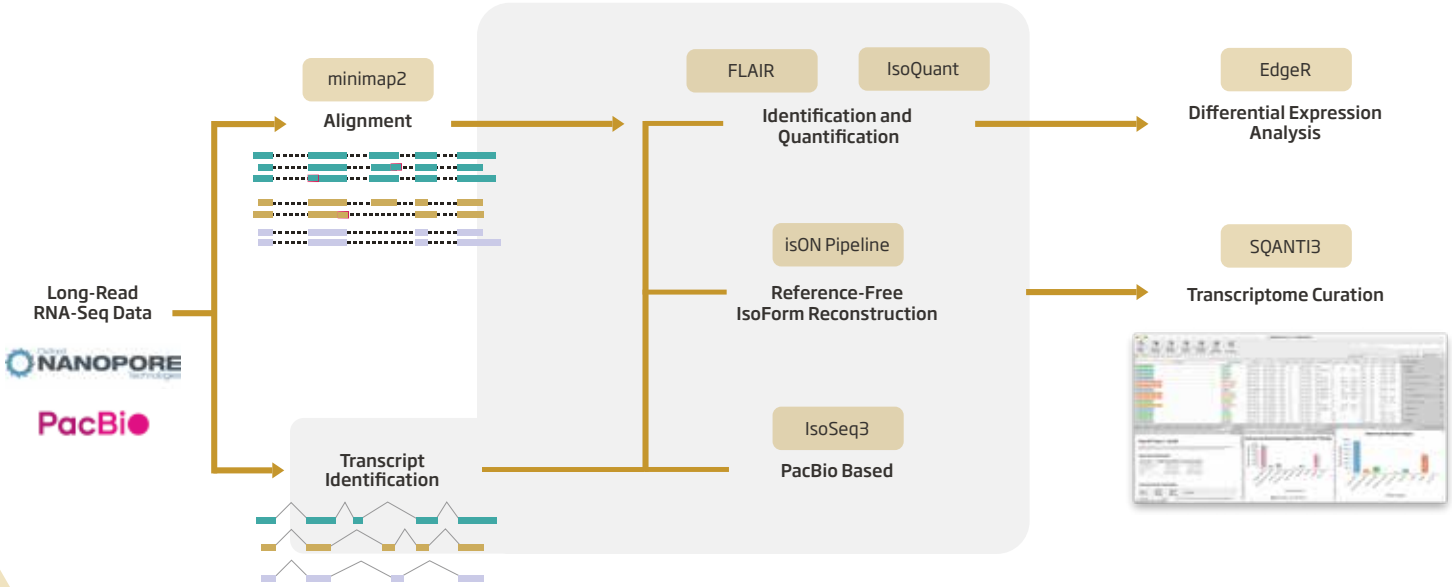
**IsoSeq**
Pre-process raw PacBio reads from different protocols with IsoSeq. OmicsBox streamlines this complicated process so you can focus on what's important.

**FLAIR**
FLAIR constructs a transcriptome from long reads and is particularly suitable for the discovery of novel transcripts. Additionally, it provides a count table for further differential expression analysis.

**IsoQuant**
IsoQuant, like FLAIR, constructs a transcriptome from long reads, however with a more reference-faithful approach. This way, only highly confident novel transcripts are reported and quantified for downstream analysis.

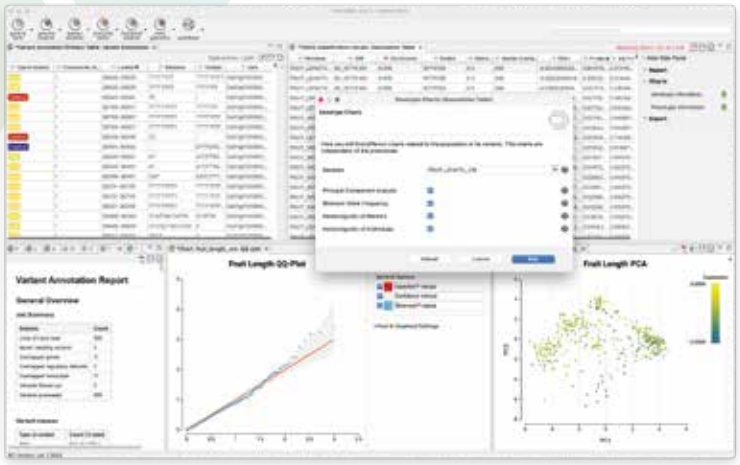
**isONpipeline**
The isONpipeline can reconstruct transcripts without requiring a reference genome. Designed for non-model species with understudied transcriptomes and genomes, this complex pipeline is made easy to use by OmicsBox.

**SQANTI3**
Once your long-read transcriptome is reconstructed, SQANTI3 assesses its quality and characterizes individual isoforms. It also filters out false-positive isoforms, providing a curated and reliable transcriptome.



# Genetic Variation

## Identify and Analyze Genetic Variations Within Populations and Species



- Variant Calling of GBS and WGS data
- Variant Filtering
- Model & Non-Model Variant Annotation
- Guided Genome-Wide Association Studies
- Population Structure Analysis

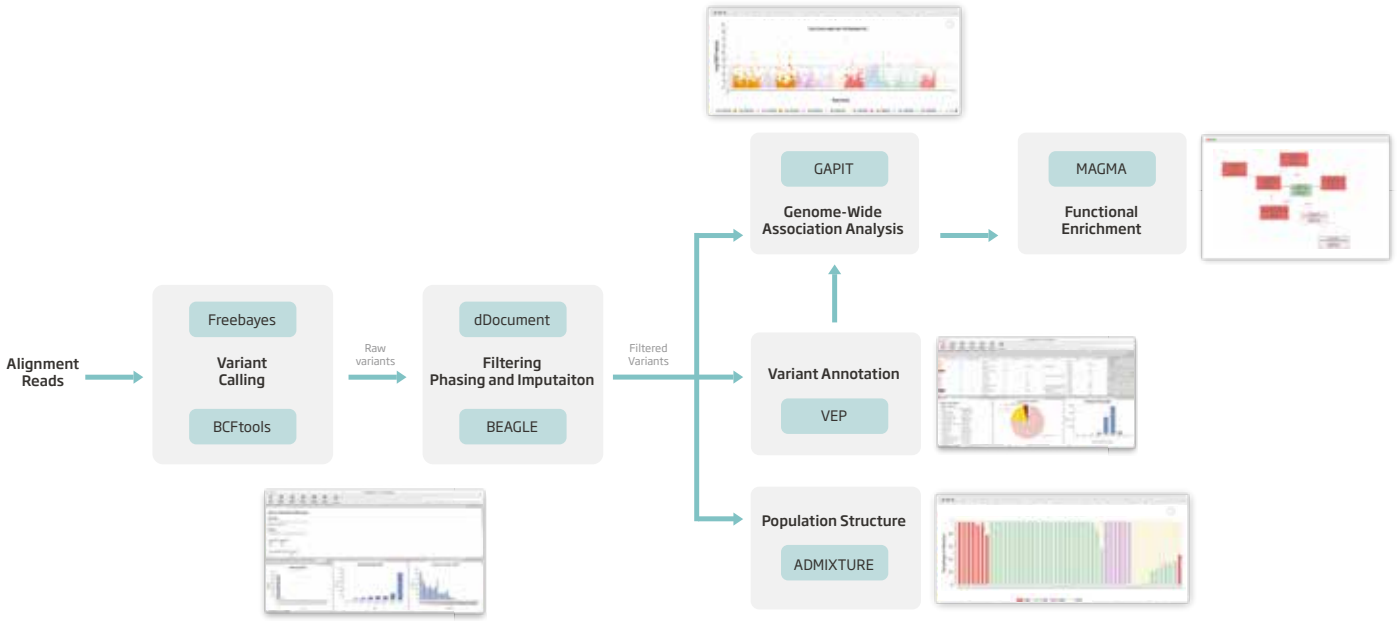
## Cutting-Edge Genetic Variation Analysis

The Genetic Variation Module enables variant calling, filtering, and annotation, as well as the association of genetic variations with specific traits or diseases through genome-wide association studies (GWAS).

The module provides two robust strategies for variant calling and filtering, utilizing the popular tools **BCFtools and FreeBayes.** Flexible options allow users to tailor the analysis to specific methodologies, such as GBS (Genotyping-by-Sequencing) or WGS (Whole Genome Sequencing), and adapt to various requirements, including ploidy level, genome coverage, and the presence of repetitive regions.

The resulting VCF files can be annotated using the Variant Effect Predictor (VEP) from Ensembl. This powerful combination of tools has demonstrated superior performance compared to alternative pipelines, as highlighted in recent review studies (e.g., Nature Sci. Rep. 12, 11331 (2022)).

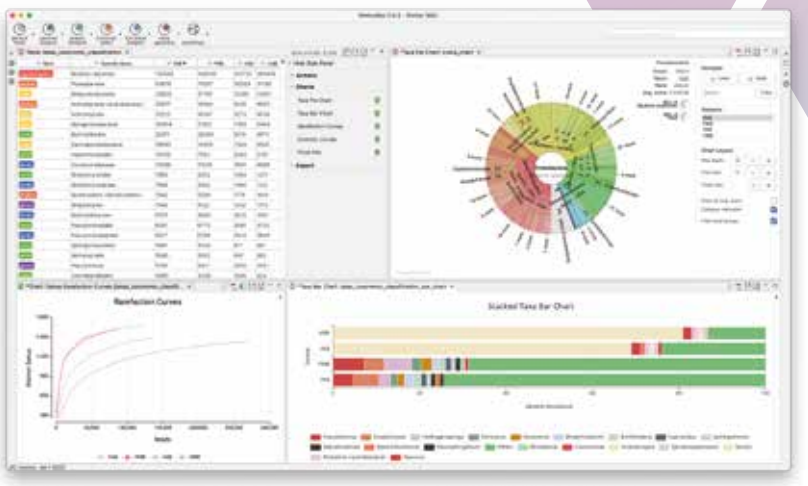
Additionally, the module facilitates guided GWAS to identify genetic variations linked to particular traits. Functional interpretation is further enhanced through **Gene Set Analysis,** providing deeper biological insights into the identified variations.



# Metagenomics

## Enables Comprehensive Microbiome Data Analysis

- Quality Control
- Metagenomic Assembly
- Taxonomic Classification
- Functional Annotation and Analysis
- Comparative Analysis



## Comprehensive WGS and 16S Microbiome Data Analysis

The Metagenomics Module provides a complete solution for microbiome data analysis, including assembly, annotation, and classification of metagenomic data. Designed to be flexible and intuitive, it integrates all essential steps for microbiome analysis and allows the creation of custom pipelines to accommodate specific analytical strategies.

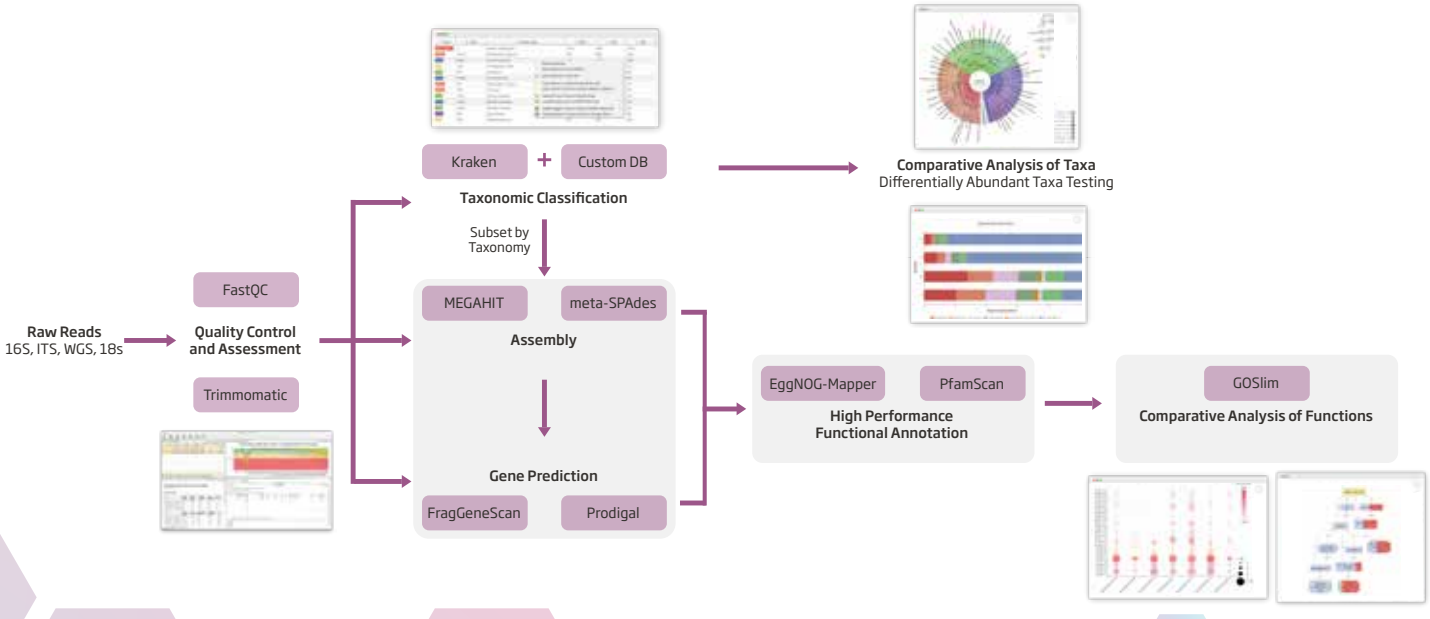
For taxonomic identification, the module leverages **Kraken2** to classify bacteria, archaea, fungi, protozoa, and viruses down to the strain level. Users can select from comprehensive databases such as RefSeq for WGS reads, or Greengenes, SILVA, GTDB, MIDAS5, and UNITE for amplicon data.

Rich visualization tools enable users to extract meaningful insights with ease.

Assembly of large datasets is supported by MetaSPAdes and MEGAHIT, offering fast and efficient processing in the cloud. Gene and protein identification can be performed using FragGeneScan for raw reads and Prodigal for assembled data.

The module also supports functional annotation through various databases, enabling deeper insights into the microbiome's biological significance. Statistical tests further allow for the identification of differential abundance in taxonomies and biological functions across samples, providing a comprehensive understanding of microbiome dynamics.

-



# Functional Analysis

## Functionally Analyze any Sequence Dataset from Scratch

- Sequence Analysis with Blast, InterPro and EggNog
- Blast2GO Functional Annotation
- GSEA and Fisher's Enrichment Analysis
- Combined Pathway Analysis with KEGG
- Rich Gene Ontology Visualizations

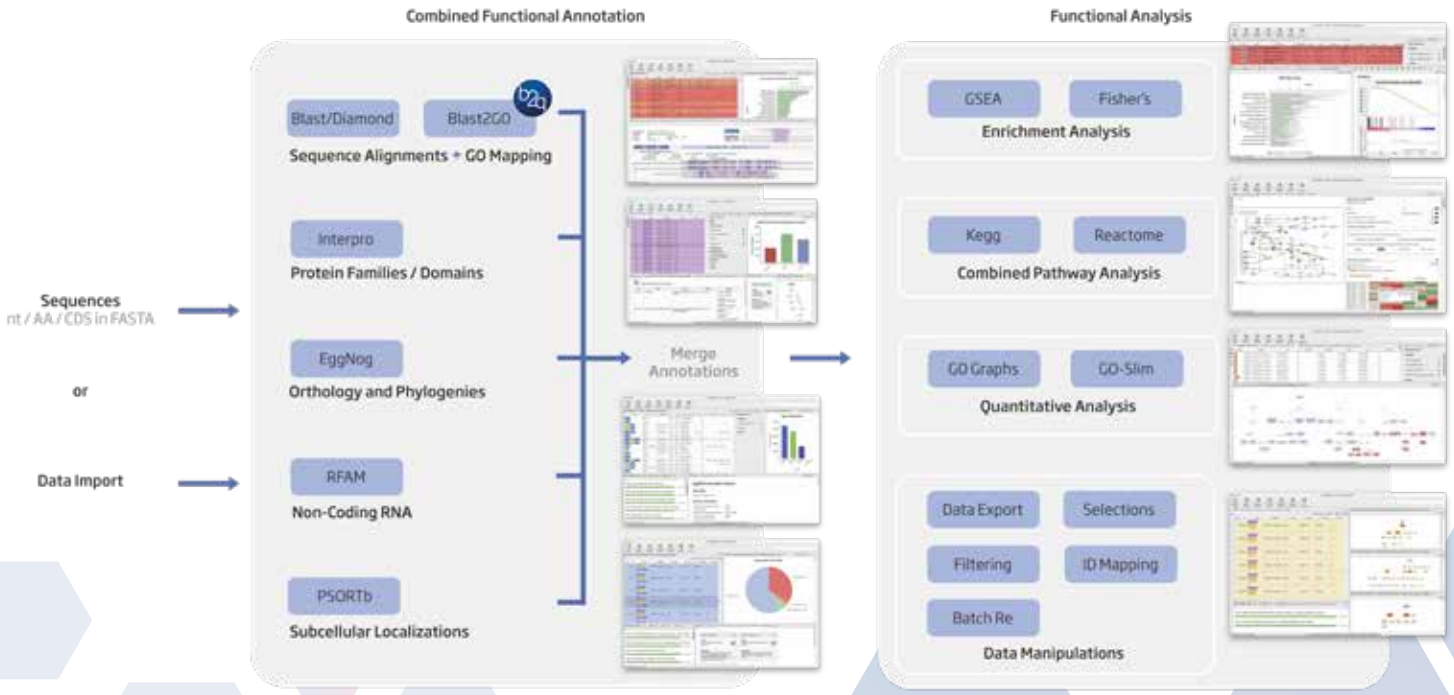
## High-Quality Functional Annotation and Enriched Pathway Analysis

The Functional Analysis module adds biological context to various data analysis approaches in Omics-Box, enhancing the interpretation of results.

The process begins with the functional annotation of datasets. The widely used **Blast2GO** methodology provides flexibility in assigning reliable functional labels to novel sequence data. This process considers annotation source quality and Gene Ontology hierarchies to ensure accuracy. Cloud-based tools like Blast, Diamond, and InterPro enable fast sequence alignment and domain searches against customizable reference datasets. Functional annotations are derived from up-to-date, well-curated databases, including UniProt and the Gene Ontology Consortium, ensuring comprehensive and reliable results.

Once the functional annotation is complete, the module offers powerful enrichment analysis tools, such as Fisher's Exact Test and GSEA, to identify over- and under-represented biological functions in the dataset.

The Combined Pathway Analysis feature takes functional insights further by identifying pathways from Reactome, Plant Reactome, and **KEGG** for any set of sequences. When combined with differential expression data, it allows pathway enrichment calculations and provides rich visualizations to help users quickly and intuitively uncover meaningful insights.

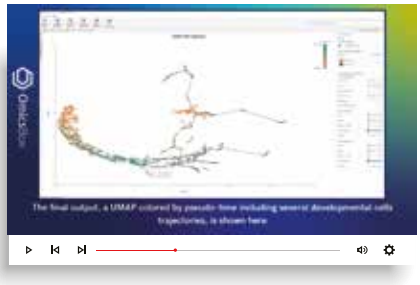


# BioBam Help Center

The BioBam Help Center is a comprehensive resource designed to support your data analysis. Whether you're new to OmicsBox or looking to refine your analysis, it offers detailed guidance, practical tips, and valuable information.

Visit the Help Center to streamline your work with OmicsBox and advance your research.

Explore our BioBam Blog for video tutorials, example analyses or the latest OmicsBox updates. Stay informed and improve your understanding of key analysis processes.



# Subscribe to OmicsBox

A subscription to OmicsBox offers a powerful solution for obtaining high-quality data analysis results for your datasets with minimal time and effort. With OmicsBox, you'll have the ability to work independently, leveraging up-to-date resources and the latest technology to analyze your data.

Additionally, you'll have access to professional support throughout your subscription, ensuring that you have the guidance and assistance you need to get the most out of your data analysis.

With the platform's user-friendly interface and intuitive tools, you'll be able to quickly and easily process and analyze your datasets, allowing you to focus on interpreting and understanding your results.

# Customize your Subscription

Tailor your subscription to meet your specific needs with flexible configuration options. Select the analysis modules you require, adjust the number of end users, and choose the subscription duration that works best for you.

1 Module
1 Year
1 User

Your
Personal
Plan

**Multiple**
Modules
& Users

biobam.com/omicsbox

www.biobam.com

BioBam is a leading bioinformatics company specializing in innovative software solutions designed to accelerate genomics research.

Committed to developing user-friendly and powerful tools, BioBam simplifies complex data analysis, enabling researchers to focus on interpreting results and uncovering new insights.

By bridging the gap between state-of-the-art bioinformatics and applied genomics research, BioBam transforms intricate analytical processes into intuitive, interactive tasks that drive scientific discovery and innovation.

biobam

BIOINFORMATICS SOLUTIONS

Developed with ♥ by BioBam