

Analyze DNA Sequencing data with InSyBio DNaseq

October 2025

Insybio Suite v3.4



InSyBio

Intelligent Systems Biology

User Manual

www.insybio.com

Introduction

DNaseq is a new tool which enables the fast and accurate pre-processing and analysis of DNA-sequencing data by non-bioinformatics experts with optimized pipelines. This tool includes the following functionalities:

- Pre-processing of DNA-sequencing data with **optimized pipelines and a user-friendly interface**
- Population analysis of genetics data for the **identification of significant genomics biomarkers**
- Integration of genomic biomarkers with InSyBio Suite's knowledge base to allow the **biological interpretation of your data**
- **Integration of genomic biomarkers with other omics biomarkers and clinical data using statistical and machine learning** functionalities of InSyBio Biomarkers.

DNA-Seq Pipeline

You can calculate the differential expression between two RNA-Seq experiments. It uses FastQC and Trimmomatic for Quality Control, HISAT2 for Alignment, FeatureCounts for Quantification and DESeq2 for Differential Expression analysis. The Rna-Seq Differential Expression we have implemented consists of 4 steps:

- A.** Quality Control using FastQC and Filtering using Trimmomatic (Optional step).
- B.** Alignment using Bowtie2, and sorting with Samtools.
- C.** Variant Calling using Freebayes.
- D.** Variant Annotation using known databases with Ensemble VEP.

Firstly, the Pipeline uses Fastqc to create a report with the input sequences quality, then trimm the sequences accordingly using Trimmomatic and create new reports with Fastqc. Then using Bowtie2 it creates the alignment SAM files with the Genome files, we sort them using SAMtools and transform them to BAM files. The BAM files are used as input of Freebayes, that creates VCF files with the variants that it detects. At the end, Variant Annotation with VEP is performed, extra information like allele frequency, SIFT variant score and the variant's id from dbSNP is annotated and some supplementary plots are created with a script using R.

We also offer a Significant Gene file creation, where if only one cohort is used we create a file with the variants with the lowest SIFT score or if multiple cohorts are used we create pairs of cohorts and calculate their significant gene variants..

To start the DNA-Seq Pipeline:

Click in the menu "InSyBio DNA-Seq" and you will be redirected to the "DNA-Seq Pipeline Dashboard", select the "Add new job" button and then:

- Select if you have Single-Cohort or Multiple-Conditions and if you have Paired or Single Ended data that you want to analyze.

Pipeline
InSyBio Beta User

What kind of data do you want to analyse?

Cohort Data: ☐ Multiple-Conditions ☒ Single-cohort

DNA-Seq Data: ☐ Paired-end ☒ Single-ended

Condition Control:

Title:

Filename:

Options

Do you want to perform initial FastQC ☐

Do you want to perform trimming?

Alignment Options

Select a reference genome: *

Specify strand information:

Filtering Options

Allele Frequency threshold value

Pipeline
InSyBio Beta User

What kind of data do you want to analyse?

Cohort Data: ☐ Multiple-Conditions ☒ Single-cohort

DNA-Seq Data: ☒ Paired-end ☐ Single-ended

Condition Control:

Title Read 1: Title Read 2:

Filename Read 1: Filename Read 2:

Options

Do you want to perform initial FastQC ☐

Do you want to perform trimming?

Alignment Options

Select a reference genome: *

Specify strand information:

Filtering Options

Allele Frequency threshold value

Pipeline

What kind of data do you want to analyse?

Cohort Data: ☒ Multiple-Conditions ☐ Single-cohort

DNA-Seq Data: ☐ Paired-end ☒ Single-ended

Condition Control: * Required information

Title:

Filename:

Condition 1:

Title:

Filename:

Options

Do you want to perform initial FastQC: ☐

Do you want to perform trimming?

Alignment Options

Select a reference genome: *

Pipeline

What kind of data do you want to analyse?

Cohort Data: ☒ Multiple-Conditions ☐ Single-cohort

DNA-Seq Data: ☒ Paired-end ☐ Single-ended

Condition Control: * Required information

Title Read 1: Title Read 2:

Filename Read 1: Filename Read 2:

Condition 1:

Title Read 1: Title Read 2:

Filename Read 1: Filename Read 2:

Options

Do you want to perform initial FastQC: ☐

Do you want to perform trimming?

Alignment Options

Select a reference genome: *

- Name Conditions/Group of files you want to Analyze.
- For each condition add single or paired files by:

- Uploading a new file of DNA-Seq Experiments in fastq format. You are redirected to the Data Store where step by step instructions guide you for both files uploading.
 - Or Selecting a file of DNA-Seq Experiments in fastq format from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.
- Select if you want to perform FastQC Quality Control to the initial Data.

Options

Do you want to perform initial FastQC ☐

Do you want to perform trimming? --Select Action--

Alignment Options

Select a reference genome: *

--Select Action--

Specify strand information:

Unstranded

Filtering Options

Allele Frequency threshold value	<input type="text" value="0.05"/>
Significant Genes threshold value	<input type="text" value="0.1"/>

DNaseq Analysis

Clear All

- Select if you want to perform trimming of the data with Trimmomatic, either with our Default Options or add your own (If trimming is selected FastQC will be performed to the trimmed data). Possible manual options are to:
 - Perform initial ILLUMINACLIP step
 - With Standard adapters (TrueSeq2, TrueSeq3 or Nextera for paired or single ended)
 - Or With Custom adapters in fasta format
 - Perform sliding window trimming
 - Drop reads below a specific length

- Cut bases off the start of a read, if below a threshold quality
- Cut bases off the end of a read, if below a threshold quality
- Cut the read to a specified length
- Cut the specified number of bases from the start of the read
- Drop the read if the average quality is below a specified value
- Trim reads adaptively, balancing read length and error rate to maximise the value of each read

Options

Do you want to perform initial FastQC
☒

Do you want to perform trimming?

YES (Set Options ▾)

Trimmomatic Options

Perform initial ILLUMINACLIP step?

YES ▾

Select standard adapter sequences or provide custom? *

Standard ▾

Adapter sequences to use: *

TruSeq3 (single-ended, f ▾)

1. Trimmomatic Operation

Sliding window trimmi ▾

Number of bases to average across:

4 ▴ ▾

Average quality required:

15 ▴ ▾

Add Trimmomatic Operation

- Select the Genome the input files belong, from our 2 built-in options (HumanGRCh38 or MouseGRCm38).

Alignment Options

Select a reference genome: *

Mouse GRCh38

Specify strand information:

Forward (FR)

Filtering Options

- Select the strandness of your input files, Unstranded, Forward or Reverse.
- Select Filtering Options, choose Allele Frequency threshold value (0.05 is recommended and the default value), and Significant Genes threshold value (0.1 is recommended and the default value)
- Last but not least select to perform the DNA-Seq Analysis.

Filtering Options

Allele Frequency threshold value 0.05

Significant Genes threshold value 0.1

DNaseq Analysis

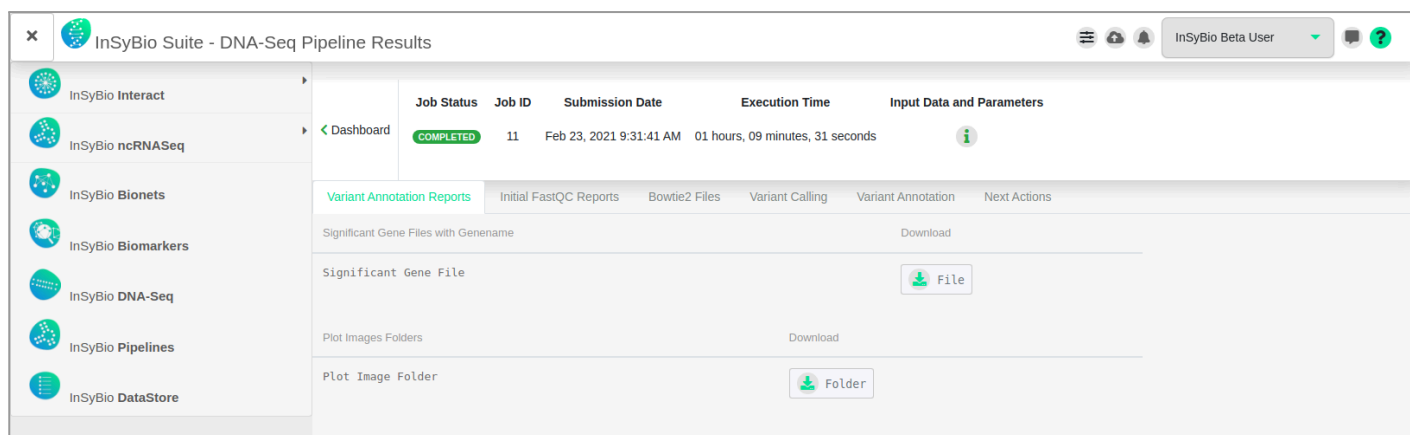
Clear All

To view the results:

By starting a calculation you are informed if it was submitted successfully. Then you can move to the DNA-Seq Pipeline and view the Dashboard, where you can view the status of your current and previous DNA-Seq Pipeline jobs.

Status	Job ID	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step	Actions
Completed	11	ERR194147 unpaired: 1. ERR194147 unpaired	2/23/21 9:31 AM	2/23/21 9:31 AM	2/23/21 10:41 AM	Plot Creation	View Results
Completed	10	SRX154486: 1. GSM947410: PrEC Cells Input gDNA-seq SRX154486 SRX154487: 2. GSM947411: LNCaP Cells Input gDNA-seq SRX154487	2/22/21 5:48 PM	2/22/21 5:48 PM	2/22/21 9:57 PM	Plot Creation	View Results
Error	8	ERR194147: 1. ERR194147 unpaired	2/19/21 3:42 PM	2/19/21 3:42 PM	2/19/21 3:43 PM	Alignment	View Details
Error	7	ERR194147: 1. ERR194147 unpaired	2/12/21 2:01 PM	2/12/21 2:01 PM	2/12/21 2:32 PM	Alignment	View Details
Completed	5	hbr: 1. HBR rep1 read1, HBR rep1 read2, 2. HBR rep2 read1, HBR rep2 read2, 3. HBR rep3 read1, HBR rep3 read2	12/18/19 7:57 AM	12/18/19 7:57 AM	12/18/19 8:27 AM	Plot Creation	View Results
Error	4	test: 1. HBR rep1 read1, HBR rep1 read2, 2. HBR rep2 read1, HBR rep2 read2	12/11/19 1:18 PM	12/11/19 1:18 PM	12/18/19 6:55 AM	Alignment	View Details
Error	3	1: 1. HBR rep1 read1, HBR rep1 read2	12/3/19 12:13 PM	12/3/19 12:13 PM	12/8/19 10:06 PM	Alignment	View Details

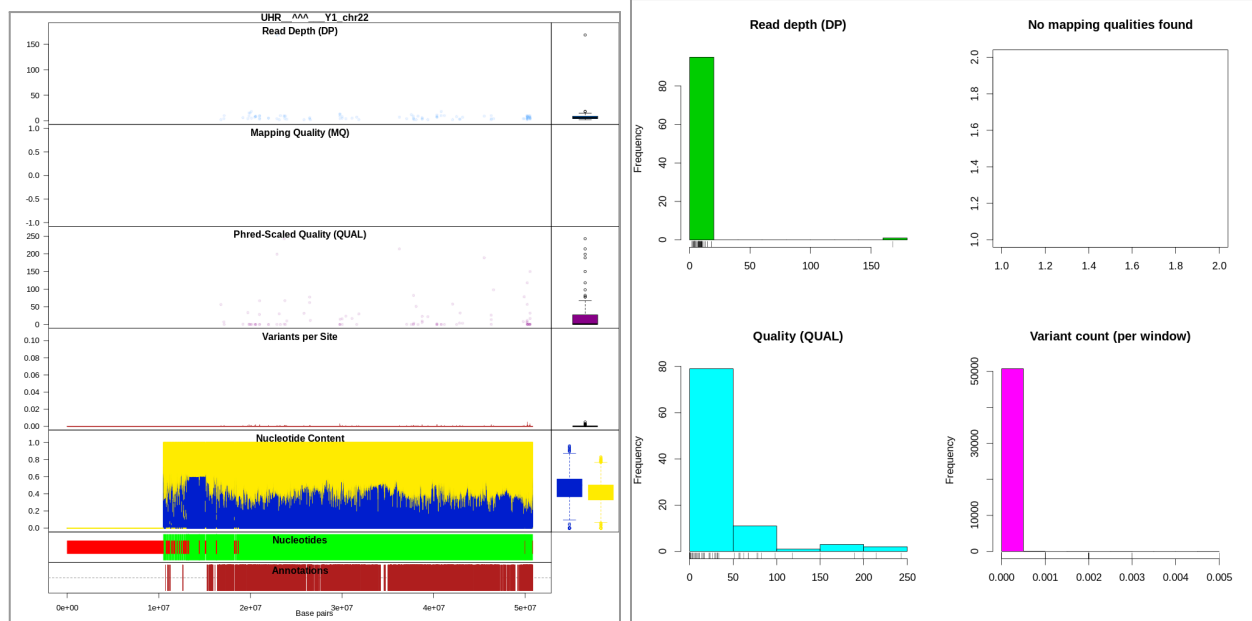
At completion of the Analysis you can select the View Results at the Actions column and view the produced files, that are separated according to the step they were produced.



In the Variant Annotations reports tab you can download visual information and the Significant Gene Files with Genename notation, and some variant alignment images.

	A	B	C	D	E
1	Ensemble Gene id	Sift score	Associated Genename	Associated Uniprot IDs	
2	ENSG00000139055	0.0	ERP27	Q96DN0,F5GYS6	
3	ENSG00000198888	0.0	MT-ND1	P03886	
4	ENSG00000198840	0.0	MT-ND3	P03897	
5					
6					
7					
8					
9					
10					

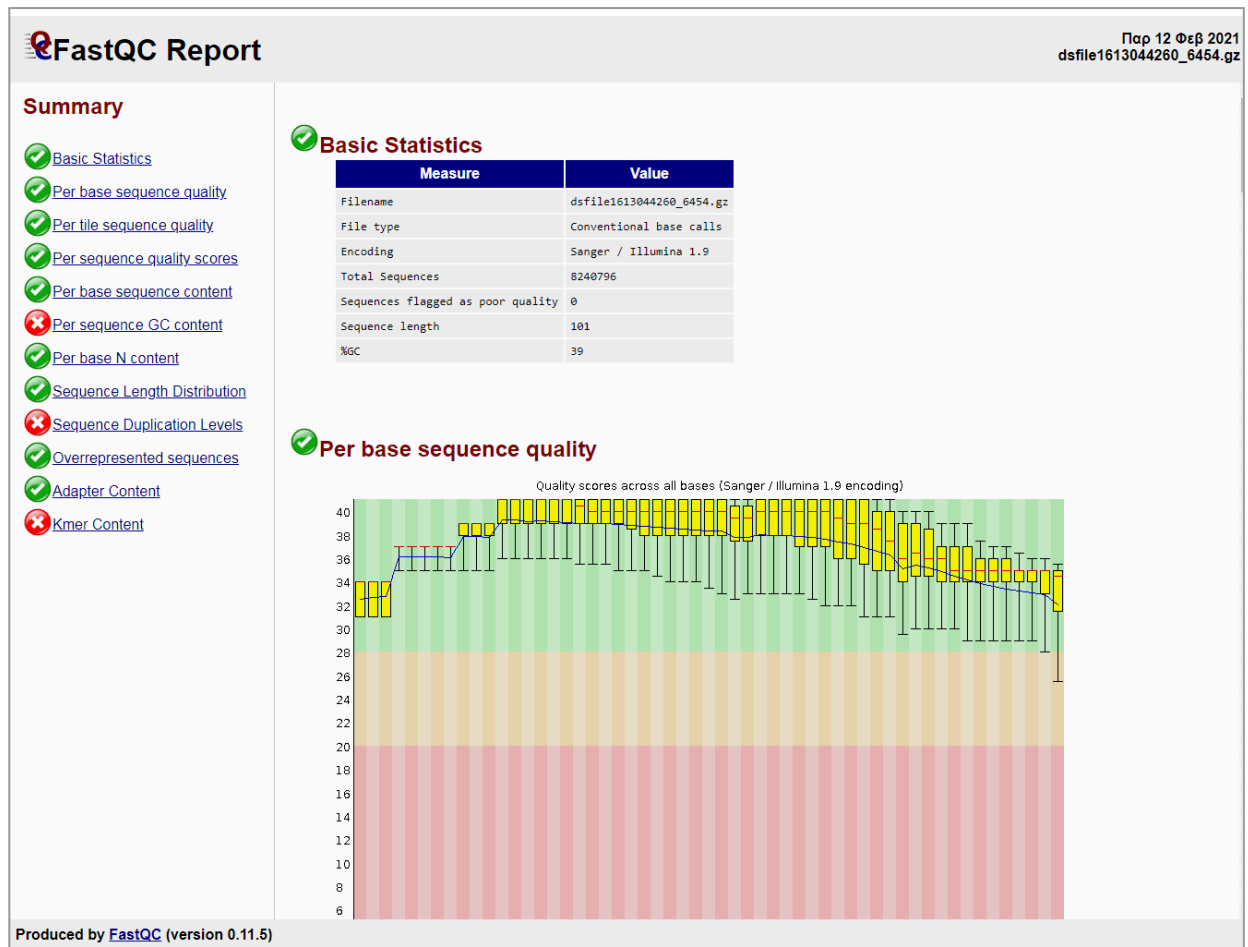
Example of the Significant Gene File being viewed with Microsoft Excel.



Example of the produced images and plots, (if there are enough data per chromosome).

The screenshot shows the InSyBio Suite - DNA-Seq Pipeline Results interface. The sidebar on the left contains navigation options: InSyBio Interact, InSyBio ncRNASeq, InSyBio Bionets, InSyBio Biomarkers, InSyBio DNA-Seq, InSyBio Pipelines, and InSyBio DataStore. The main panel displays job status (COMPLETED), job ID (11), submission date (Feb 23, 2021 9:31:41 AM), execution time (01 hours, 09 minutes, 31 seconds), and input data and parameters. It also shows a list of reports including Variant Annotation Reports, Initial FastQC Reports, Bowtie2 Files, Variant Calling, Variant Annotation, and Next Actions. A FastQC Report is selected, showing a download button and a view HTML page button. The input data is listed as 'DNASeq Job-11 Fastqc zip file ERR194147_unpaired rep1'.

If Initial FastQC is selected, in the Initial FastQC reports the FastQC reports of the input files can be downloaded.



Example of a FastQC Report html file, one for each experiment is produced.

InSyBio Suite - DNA-Seq Pipeline Results InSyBio Beta User

Job Status **Job ID** **Submission Date** **Execution Time** **Input Data and Parameters**

COMPLETED 5 Dec 18, 2019 7:57:02 AM 00 hours, 30 minutes, 02 seconds

Variant Annotation Reports **Trimmed FASTQ Files** **Trimmed FastQC Reports** **Bowtie2 Files** **Variant Calling** **Variant Annotation** **Next Actions**

Trimmed FASTQ File **Download**

DNaseq Job-5 trimmend paired file of hbr rep1 read1 (dsfile1557128487_9359_trimmed.gz); [File](#)

DNaseq Job-5 trimmend paired file of hbr rep1 read2 (dsfile1557128516_9128_trimmed.gz); [File](#)

DNaseq Job-5 trimmend paired file of hbr rep2 read1 (dsfile1557128550_6204_trimmed.gz); [File](#)

DNaseq Job-5 trimmend paired file of hbr rep2 read2 (dsfile1557128587_1781_trimmed.gz); [File](#)

In the Trimmed FASTQ Files, the output Fastq files after trimming can be downloaded.

The screenshot shows the InSyBio Suite - DNA-Seq Pipeline Results interface. On the left is a sidebar with navigation links: InSyBio Interact, InSyBio ncRNASeq, InSyBio Bionets, InSyBio Biomarkers, InSyBio DNA-Seq, InSyBio Pipelines, and InSyBio DataStore. The main content area displays job details for a completed job (Job ID: 5, Submission Date: Dec 18, 2019 7:57:02 AM, Execution Time: 00 hours, 30 minutes, 02 seconds). Below the job details, there are tabs for Variant Annotation Reports, Trimmed FASTQ Files, Trimmed FastQC Reports (selected), Bowtie2 Files, Variant Calling, Variant Annotation, and Next Actions. Under the Trimmed FastQC Reports tab, there are three entries, each with a 'Download' button and a 'View Html Page' button. The entries are: s:58:"DNaseq Job-5 after trimming Fastqc zip file hbr repl read1";, s:58:"DNaseq Job-5 after trimming Fastqc zip file hbr repl read2";, and s:58:"DNaseq Job-5 after trimming Fastqc zip file hbr rep2 read1";.

In the Trimmed FastQC reports the FastQC reports of the trimmed files can be downloaded.

The screenshot shows the InSyBio Suite - DNA-Seq Pipeline Results interface. On the left is a sidebar with navigation links: InSyBio Interact, InSyBio ncRNASeq, InSyBio Bionets, InSyBio Biomarkers, InSyBio DNA-Seq, InSyBio Pipelines, and InSyBio DataStore. The main content area displays job details for a completed job (Job ID: 18, Submission Date: Feb 11, 2021 12:19:48 PM, Execution Time: 00 hours, 32 minutes, 31 seconds). Below the job details, there are tabs for Variant Annotation Reports, Bowtie2 Files (selected), Variant Calling, Variant Annotation, and Next Actions. Under the Bowtie2 Files tab, there are three entries, each with a 'Download' button and a 'File' button. The entries are: SAM File (DNaseq Job-18 Bowtie2 alignment file err194147_1.sam (err194147_1.sam);), BAM File (DNaseq Job-18 BAM file err194147_1.bam (err194147_1.bam);), and Run Info (Alignment Info). The Alignment Info entry has a 'Download' button and a 'File' button.

In the Bowtie2 files tab, the Bowtie2 alignment sam and bam files can be downloaded.

Example of Alignment information inside the bowtie2_report.txt:

```
8131633 reads; of these:

    8131633 (100.00%) were unpaired; of these:

    34333 (0.42%) aligned 0 times

    4183088 (51.44%) aligned exactly 1 time

    3914212 (48.14%) aligned >1 times

99.58% overall alignment rate
```

Job Status	Job ID	Submission Date	Execution Time	Input Data and Parameters
COMPLETED	5	Dec 18, 2019 7:57:02 AM	00 hours, 30 minutes, 02 seconds	

Variant Annotation Reports	Trimmed FASTQ Files	Trimmed FastQC Reports	Bowtie2 Files	Variant Calling	Variant Annotation	Next Actions
Variant Call Files						
DNaseq Job-5 Variant Annotation file (hbr_1.vcf);				Download		
				File		
DNaseq Job-5 Variant Annotation file (hbr_2.vcf);				File		
DNaseq Job-5 Variant Annotation file (hbr_3.vcf);				File		

In the Variant Calling tab the unfiltered VCF file is provided as created by Freebayes and is available to be downloaded.

The screenshot displays the InSyBio Suite - DNA-Seq Pipeline Results interface. The left sidebar contains navigation links: InSyBio Interact, InSyBio ncRNASeq, InSyBio Bionets, InSyBio Biomarkers, InSyBio DNA-Seq, InSyBio Pipelines, and InSyBio DataStore. The main panel shows a table with columns: Job Status, Job ID, Submission Date, Execution Time, and Input Data and Parameters. The Job Status is 'COMPLETED', Job ID is '5', Submission Date is 'Dec 18, 2019 7:57:02 AM', and Execution Time is '00 hours, 30 minutes, 02 seconds'. Below the table, there are tabs for Variant Annotation Reports, Trimmed FASTQ Files, Trimmed FastQC Reports, Bowtie2 Files, Variant Calling, Variant Annotation (selected), and Next Actions. The Variant Annotation tab shows a list of files for download, including Missense Variant Vep Files, DNaseq Job-5 Filtered missense_variants Variant Annotation file (hbr_1_missense_annotations.vcf), DNaseq Job-5 Filtered missense_variants Variant Annotation file (hbr_2_missense_annotations.vcf), DNaseq Job-5 Filtered missense_variants Variant Annotation file (hbr_3_missense_annotations.vcf), Protein Altering Variants, DNaseq Job-5 Filtered protein_altering_variants and AF < 0.05 Variant Annotation file (hbr_1_filtered_annotations.vcf), DNaseq Job-5 Filtered protein_altering_variants and AF < 0.05 Variant Annotation file (hbr_2_filtered_annotations.vcf), DNaseq Job-5 Filtered protein_altering_variants and AF < 0.05 Variant Annotation file (hbr_3_filtered_annotations.vcf), All Variants, DNaseq Job-5 Variant Annotation file (hbr_1_annotations.vcf), and DNaseq Job-5 Variant Annotation file (hbr_2_annotations.vcf). Each file has a 'Download' button.

In the Variant Annotation tab the different Annotated Variant vcf files for each sample can be downloaded. Missense Variant Vep files, Protein Altering Variants and All Variants are available.

The screenshot displays the InSyBio Suite - DNA-Seq Pipeline Results interface, showing the Next Actions tab. The left sidebar is the same as the previous screenshot. The main panel shows a table with columns: Job Status, Job ID, Submission Date, Execution Time, and Input Data and Parameters. The Job Status is 'COMPLETED', Job ID is '5', Submission Date is 'Dec 18, 2019 7:57:02 AM', and Execution Time is '00 hours, 30 minutes, 02 seconds'. Below the table, there are tabs for Variant Annotation Reports, Trimmed FASTQ Files, Trimmed FastQC Reports, Bowtie2 Files, Variant Calling, Variant Annotation, and Next Actions (selected). The Next Actions tab shows a section titled 'Continue your Analysis in InSyBio Suite'. It lists 'Significant Gene Files' and provides a 'Download' button. Below this, there is a table with columns: Significant Gene File, Download, and Next Action. The table contains one row: 'Significant Gene File - Condition 1', 'Download', and a dropdown menu with the text '--Select Action--'.

In the Next Action tab, Significant Genes files, with the provided threshold (default 10%) the most significant genes, for each cohort are provided. They can be downloaded or used as input in **InSyBio Interact**, to **Create Networks** from that set of significant genes based on the protein-protein interactions knowledge base of

InSyBio Interact, or to perform GO Term **Enrichment Analysis** from that set of biomarkers based on the protein-go term correlation knowledge base of InSyBio Interact..

Cloud computing Infrastructurer and Security Certifications

InSyBio Suite and all its tools are running over the cloud computing as a service infrastructure of Vultr (<https://www.vultr.com>), at the Amsterdam (Netherlands) facilities, offering the following security attestations and certifications (SOC 2+ (HIPAA), PCI (Merchant), CSA Star Level 1, ISO/IEC 20000-1:2018, ISO/IEC 27001:2022, ISO/IEC 27017:2015, ISO/IEC 27018:2019).

How to get InSyBio Interact

To request a free one month license of InSyBio Suite please email us at info@insybio.com.

To purchase InSyBio Interact commercial version 3.3 please contact us at sales@insybio.com.

About Us

InSyBio Inc is a bioinformatics pioneer company (www.insybio.com) in personalized healthcare, that focuses on developing computational frameworks and tools for the analysis of complex life-science and biological data in order to develop

predictive integrated biomarkers (biomarkers of various categories) with increased prognostic and diagnostic aspects for the personalized Healthcare Industry.

InSyBio Suite consists of tools for providing integrated biological information from various sources, while at the same time it is empowered with robust, user-friendly and installation-free bioinformatics tools based on intelligent algorithms and methods.

COPYRIGHT NOTICE

External Publication of InSyBio Inc - Any InSyBio information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the InSyBio Inc. A draft of the proposed document should accompany any such request. InSyBio Inc reserves the right to deny approval of external usage for any reason.

Copyright 2025 InSyBio Inc. Reproduction without written permission is completely forbidden.