# Analyze DNA Sequencing data with InSyBio DNASeq

### March 2022

#### Insybio Suite v3.0



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.

Pineline		🚍 🚳 🌒 InSyBio Beta Use
ripenne		
What kind of data do you want to analyse?		
Cohart Data:	short	
Sonore Bata. O Multiple-conditions of Single-co	and c	
DNA-Seq Data: O Paired-end O Single-ended		
Title: FPP194147	ă	
Filename: * dsfile1613044260_6454.gz		
Select from Data Store     OUpload to Data Store	e	
	-	
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:		
unstranged	Ť	
Filtering Options		
Allolo Fraguency threshold value		

Pipeline		🚔 🚳 🌲 🛛 InSyBio Beta User 🔍 🛒 🍞
What kind of data do you want to analyse?		Θ
Cohort Data: O Multiple-Conditions Single-cohort		
DNA-Seq Data: • Paired-end Single-ended		
Condition Control: *ERR194147	۵.	* Required information
Title Read 1: * ERR194147 read1	Title Read 2: * ERR194147 read2	
Filename Read 1: * dsfile1613032895_3036.gz	Filename Read 2: * dsfile1613033004_5428.gz	
Select from Data Store     OUpload to Data Store	Select from Data Store Oupload to Data Store	
		Add Pair
		Clear All Files
Options		
Do you want to perform initial FastQC		
Do you want to perform trimming?	Select Action 🗢	
Alignment Options		
Select a reference genome: *		
- Salact Action		
- Defect Action-	•	
Specify strand information:		
Unstranded	÷	
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: *err194147	۵. ا	* Required information
Title: * EPP194147 unpaired		
Hiename: dsmle1613044260_6454.gz		
Select from Data Store		Add Ella
		Aurite
Condition 1: *		
Filename: *		
Select from Data Store		
Select non Data Store		Add File
		Audine
Add Condition		Clear All Files
Ontions		
Options		
Do you want to perform initial FastQC		
Do you want to perform trimming?	Select Action 💠	
Alignment Options		
Select a reference genome: *		
Pipeline		🚍 🚳 🌲 🛛 InSyBio Beta User 🛛 💻 🍞
Pipeline		🚍 🚳 🌒 InSyBio Beta User 💌 💻 😨
Pipeline What kind of data do you want to analyse?		🚍 💿 🌒 InSyBio Beta User 👻 💻 🍞
Pipeline What kind of data do you want to analyse? Cohort Data: • Multiple-Conditions · Single-cohort		🚍 🍙 🌒 InSyBio Beta User 👻 🛒 🍞
Pipeline What kind of data do you want to analyse? Cohort Data: • Multiple-Conditions · Single-cohort		🚍 🙆 🌒 InSyBio Beta User 👻 💻 😭
Pipeline What kind of data do you want to analyse? Cohort Data: • Multiple-Conditions · Single-cohort DNA-Seq Data: • Paired-end · Single-ended		🛱 🙆 🌒 InSyBio Beta User 👻 💭 💭 🚍
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: +[ERR194147		InsyBio Beta User
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: +[ERR194147 Title Read 1: + [ERR194147 read1 Filter Paired 1: + [ERR194147 read1	Title Read 2: * ERR194147 read2	InsyBio Beta User Figure 1 Required information
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:       #ERR194147         Title Read 1:       * ERR194147 read1         Filename Read 1:       * defile1613032895_3036.gz	Title Read 2: * ERR194147 read2 Filename Read 2: * dsfile1613033004_5428.gz	mSyBio Beta User The system of the
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:       •ER194147         Title Read 1:       •ER194147 read1         Filename Read 1:       • disfiel 613032895_3036.gz         Image: Solid Control State Store       Image: Outplace Store         Image: Solid Control State Store       Image: Outplace Store	Title Read 2: * ER1194147 read2 Filename Read 2: * dsfile1613033004_5428.gz	InSyBio Beta User
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: •[ER194147 Title Read 1: •[ER194147 read1 Filename Read 1: • disfile1613032895_3036.gz Control of the Paired Store Upload to Data Store	Title Read 2: * ER1194147 read2 Filename Read 2: * dsfile1613033004_5428.gz	InsyBio Beta User       Image: SyBio Be
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: •[ER194147 Title Read 1: •[ER194147 Title	Title Read 2: * ER194147 read2 Filename Read 2: * dsfile1613033004_5428.gz	mSyBio Beta User       ImSyBio Beta User     ImSyBio Beta User       Imscription       Imscription       * Required information       Add Pair
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: •[ER194147 Title Read 1: • ©ER194147 read1 Filename Read 1: • ©ER194147 Title Read 1: • ©ER19447 Title Read 1: •	Title Read 2:	E So () InSyBio Beta User
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         Condition Control:       •[ER194147         Title Read 1:       • ER194147         Filename Read 1:       • disfiel 613032895_3036.gz         Condition 1:       • Enter Home Data Store         Condition 1:       • Enter Home Data Store         Filename Read 1:       • Enter Home Data Store	Title Read 2: * dsfile1613033004_5428.gz Title Read 2: * dsfile1613033004_5428.gz Title Read 2: * Title Read 2: * Filename Read 2: *	InsyBio Beta User
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Condition Control:       •[ER194147         Title Read 1:       • ER194147         Filename Read 1:       • Gsfiel1613032895_3036.gz         Condition 1:       • Gsfiel1613032895_3036.gz         Condition 1:       • Gsfiel1613032895_3036.gz         Condition 1:       • Title Read 1:         Filename Read 1:       • Upload to Data Store	Title Read 2: * GSR194147 read2 Filename Read 2: * GSR194147 read2 Filename Read 2: * GSR194147 read2 Title Read 2: * Upload to Data Store Title Read 2: * Filename Read 2: * Filename Read 2: *	E So So Insylio Beta User
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         Condition Control:       •[ER194147 read1         Filename Read 1:       • Gsfield 10302895_3036.gz         Condition 1:       • Gsfield 10302895_3036.gz         Condition 1:       • Gsfield 10302895_3036.gz         Condition 1:       • Title Read 1:         Filename Read 1:       •         Filename Read 1:       •         Sinder from Data Store       •         Onder three Data Store       •         Onder three Data Store       •	Title Read 2: * dsfile1613033004_5428.gz Filename Read 2: * dsfile1613033004_5428.gz Title Read 2: * Title Read 2: * Filename Read 2: * Filename Read 2: * Filename Read 2: *	E So (myblo Beta User ) (g) (g) • Required information Add Pair Add Pair
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Condition Control:       •[ER194147         Title Read 1:       • [ER194147 read]         Filename Read 1:       • [ER194147 read]         Title Read 1:       • [ER194161302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Add Condition 1:       • [ER194161302895_3036.gz	Title Read 2: * Gsfile1613033004_5428.gz         Filename Read 2: * Gsfile1613033004_5428.gz         Title Read 2: * Upload to Data Store         Title Read 2: *         Filename Read 2: *	E So Insysto Beta User Required information Add Pair Add Pair
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Condition Control:       •[ER194147         Title Read 1:       • [ER194147 read1         Filename Read 1:       • [Gsfiel161302895_3036.gz         Condition 1:       • [Gsfiel161302895_3036.gz         Condition 1:       • [Gsfiel161302895_3036.gz         Condition 1:       • [Gsfiel161302895_3036.gz         Condition 1:       • [Gsfiel161302895_3036.gz         Add Condition       • [Gsfiel161302895_3036.gz	Title Read 2:  Select from Data Store Select from Select from Select Store S	E So So Insybio Beta User
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:       •[ER194147         Title Read 1:       • ER194147 read1         Filename Read 1:       • diffici 61302895_3036.gz         Condition 1:       • Image: Condition 1:         Title Read 1:       • Upload to Data Store         Condition 1:       • Image: Condition 1:         Milename Read 1:       • Image: Condition 1:         Options       • Image: Condition 1:	Title Read 2: • GRF194147 read2 Filename Read 2: • dsfile1613033004_5428.gz Filename Read 2: • Upload to Data Store Title Read 2: • Filename Read 2: • Filename Read 2: • Filename Read 2: • Filename Read 2: •	InSyBio Beta User          Image: Second se
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Single-cohort         DNA-Seq Data:       Paired-end         Condition Control:       •[ER194147         Title Read 1:       • ER194147 read1         Filename Read 1:       • disfiel 61302895_3036.gz         Condition 1:       • If Condition 1:         Title Read 1:       • Upload to Data Store         Condition 1:       • If Condition 1:         Material       • If Condition 1:         Add Condition       • If Condition         Options       Do you want to perform initial FastQC	Title Read 2:	E So So Source Set User Required information Add Pair Add Pair Clear All Files
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:  Paired-end Single-ended Condition Control:  Paired-end Single-ended Condition Control:  Paired Paired-end Single-ended Condition Control:  Paired Paired-end Single-ended Condition 1:  Paired Paired Single Single-ended Condition Options Do you want to perform initial FastQC Do you want to perform timming?	Title Read 2: • dsfile101033004_5428.gz Filename Read 2: • dsfile101033004_5428.gz Title Read 2: • • • • • • • • • • • • • • • • • •	E So Source User Required information Add Pair Add Pair Clear All Files
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end         Condition Control:       •[ER194147         Title Read 1:       • [ER194147 read1         Filename Read 1:       • [ER1941610302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Mide Read 1:       • [ER194161302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Condition 1:       • [ER194161302895_3036.gz         Add Condition       • [ER194161302895_3036.gz         Options       Do you want to perform initial FastQC         Do you want to perform trimming?       • [ER194147	Title Read 2: • GRI194147 read2 Filename Read 2: • dsfile1613033004_5428.gz Title Read 2: • Upload to Data Store Title Read 2: • Filename Read	Equired information • Required information Add Pair Clear All Files
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end       Single-ended         Condition Control:       •[ER194147 read1         Title Read 1:       • [ER194147 read1         Filename Read 1:       • [ER1941613032895_3036.gz         Condition 1:       • [ER1941613032895_3036.gz         Condition 1:       • [ER19417 read1         Title Read 1:       • [ER1941613032895_3036.gz         Condition 1:       • [ER1941613032895_3036.gz         Object them Data Store       • [Dipload to Data Store]         Add Condition       • [ER19416130368]         Options       Do you want to perform initial FastQC         Do you want to perform trimming?       Alignment Options	Title Read 2: • GRI194147 read2 Filename Read 2: • dsfile1613033004_5428.gz  Filename Read 2: • Upload to Data Store  Title Read 2: • Filename Read Read Read Read Read Read Read Rea	E So () InSyBio Beta User ) () () () • Required information • Required information Add Pair Add Pair Clear All Files
Pipeline         What kind of data do you want to analyse?         Cohort Data:       Multiple-Conditions         DNA-Seq Data:       Paired-end       Single-cohort         DNA-Seq Data:       Paired-end       Single-ended         Condition Control: •[ER194147 read1       Filename Read 1: • •[ER194147 read1         Filename Read 1: • •[ER194147 read2       Upload to Data Store         Condition 1: •[Er1941613022095_3036.gz       Options         Condition       1: •[Er1941613022095_3036.gz         Condition 1: •[Er1941613022095_3036.gz       Options         Condition       1: •[Er1941613022095_3036.gz         Condition 1: •[Er1941613022095_3036.gz       Options         Options       Options         Do you want to perform initial FastQC       Do you want to perform trimming?         Alignment Options       Alignment Options	Title Read 2:  Generative form but a store Title Read 2:  Generative form but a store Title Read 2:  Filename Read Read Read Read Read Read Read Rea	E Systio Beta User   Insystio Beta User     • Required information     • Required information     Add Pair     Clear All Files

- 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 Fas	tQC				
	Do you want to perform trimming	?	Select /	Action	\$	
	Alignment Options					
	Select a reference genc	ime: *				
	Select Action		\$			
	Specify strand informat	ion:				
	Unstranded		\$			
	Filtering Options					
	Allele Frequency threshold value	0.05				
	Significant Genes threshold 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	Ø	
Do you want to perform trimming?	YES (Set Options 🗢	
Trimmomatic Options		
Perform initial ILLUMINACLIP step?	YES 💠	
Select standard adapter sequences o provide custom? *	r Standard \$	
Adapter sequences to use: *	TruSeq3 (single-ended, f	
1. Trimmomatic Operation Sliding window trimmi ◆		
average across:		
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 GRCm38	÷
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.

×	💮 InSyBio Suite - DNA-Seq F	Pipeline Das	hboard			ŧ	InSyB	io Beta User	- 🖲 😮
	InSyBio Interact	🔁 Add new Jo	b			<b>T</b> Filter Jobs	Show All -	3 0 Ipleted Running	0 6 Pending Error
	InSyBio Biomarkers	Status	Job ID 14	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step 11	Actions 11
	InSyBio Pipelines	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
	InSyBio DataStore	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 repl readl, HBR repl read2, 2. HBR rep2 readl, 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.

×	💮 InSyBio Suite - DNA-Seq P	nSyBio Suite - DNA-Seq Pipeline Results							
	InSyBio Interact	Job Status	Job ID Submission Date	Execution Time Input Data and Parameters					
	InSyBio ncRNASeq	< Dashboard COMPLETED	11 Feb 23, 2021 9:31:41 AM 01 hours	s, 09 minutes, 31 seconds					
	InSyBio Bionets	Variant Annotation Reports	Initial FastQC Reports Bowtie2 Files	Variant Calling Variant Annotation Next Action	IS				
	InSyBio Biomarkers	Significant Gene Files with Gene	ename	Download					
	InSyBio DNA-Seq	Significant Gene File		🛃 File					
	InSyBio Pipelines	Plot Images Folders		Download					
	InSyBio DataStore	Plot Image Folder		🛃 Folder					

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

	Αυτόματη αποθήκευση 🤇	• دو 🖪 🗩	~~	Είσοδος 🖬	– 🗆 X
A	οχείο <b>Κεντρι</b> Εισαγα	ω Διάταξr Τύποι	Δεδομέ Αναθεώ Προβολ	Βοήθεις 🔎 Ανσ	ιζήτη 🖻 🖵
Πρ	Δ όχειρο Γραμματοσειρά	Ξ % Στοίχιση Αριθμόσ	<ul> <li>Μορφοποίηση υπό</li> <li>Μορφοποίηση ως τ</li> <li>Στυλ κελιών ~</li> </ul>	όρους ~ 📺 τίνακα ~ Κελιά Β	Σπεξεργασία ~
C1	.0 - : ×	$\sqrt{f_x}$			v
	А	В	с	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					
10				1	
-1-1	err1941	17_fastqcsignificant	_gen 🕂 : [	•	
Έτο	ιμο			▣ 끤 – ───	+ 100%

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).

×	💮 InSyBio Suite - DNA-Seq F	InSyBio Suite - DNA-Seq Pipeline Results 🚔 🙆 🗎 InSyBio Beta User 🔹 🗏 🗘				
	InSyBio Interact	Job Status Job ID Submission Date Execution Time Input Data and Parameters				
	InSyBio ncRNASeq	< Dashboard COMPLETED 11 Feb 23, 2021 9:31:41 AM 01 hours, 09 minutes, 31 seconds				
	InSyBio Bionets	Variant Annotation Reports Initial FastQC Reports Bowtie2 Files Variant Calling Variant Annotation Next Actions				
	InSyBio Biomarkers	FastQC Report Download View Html Page				
	InSyBio DNA-Seq	DNASeq Job-11 Fastqc zip file ERR194147_unpaired rep1				
	InSyBio Pipelines					
	InSyBio DataStore					

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 🔹 💭	?
InSyBio Interact	Job Status Job ID Submission Date Execution Time Input Data and Parameters		
InSyBio ncRNASeq	COMPLETED 5 Dec 18, 2019 7:57:02 AM 00 hours, 30 minutes, 02 seconds		
InSyBio Bionets	Variant Annotation Reports Trimmed FASTQ Files Trimmed FastQC Reports Bowtie2 Files Variant Calling Variant Annotation Net	ext Actions	
InSyBio Biomarkers	Trimmed FASTQ File Download		
InSyBio DNA-Seq	DNASeq Job-5 trimmend paired file of hbr repl readl (dsfile1557128487_9359_trimmed.gz);		
InSyBio Pipelines	DNASeq Job-5 trimmend paired file of hbr repl read2 (dsfile1557128516_9128_trimmed.gz);		
InSyBio DataStore	DNASeq Job-5 trimmend paired file of hbr rep2 read1 (dsfile1557128550_6204_trimmed.gz);		
	DNASeq Job-5 trimmend paired file of hbr rep2 read2 (dsfile1557128587_1781_trimmed.gz);		

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

	<b>49</b>	
×	💮 InSyBio Suite - DNA-Seq P	Pipeline Results 🚔 💩 🌡 InSyBio Beta User 🔻 🛡 💡
	InSyBio Interact	Job Status Job ID Submission Date Execution Time Input Data and Parameters
	InSyBio ncRNASeq	COMPLETED         5         Dec 18, 2019 7:57:02 AM         00 hours, 30 minutes, 02 seconds         1
	InSyBio <b>Bionets</b>	Variant Annotation Reports Trimmed FASTQ Files Trimmed FastQC Reports Bowrie2 Files Variant Calling Variant Annotation Next Actions
	InSyBio <b>Biomarkers</b>	Trimmed FastQC Report Download View Html Page
	InSyBio <b>DNA-Seq</b>	s:58:"DNASeq Job-5 after trimming Fastqc zip file hbr repl read1"; File dsfile1557128487_9359_trimmed_fastqc
	InSyBio <b>Pipelines</b>	s:58:"DNASeq Job-5 after trimming Fastqc zip file hbr rep1 read2"; File dsfile1557128516_9128_trimmed_fastqc
	InSyBio DataStore	s:58:"DNASeq Job-5 after trimming Fastqc zip file hbr rep2 read1"; File dsfile1557128550_6204_trimmed_fastqc

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

q	Pipeline Results							
	< Dashboard	Job Status	<b>Job ID</b> 18	Subr Feb 11, 2	<b>nission Date</b> 2021 12:19:48 PM	Execution Time 00 hours, 32 minutes, 31 seconds	Input Data and Parameters	5
	Variant Anno	tation Reports	Bowtie	e2 Files	Variant Calling	Variant Annotation Next Ac	tions	
	SAM File							Download
	DNASeq Job-1	8 Bowtie2 alio	gnment fil	le err194	147_1.sam (err194	4147_1.sam);		🛓 File
	BAM File						Download	
	DNASeq Job-1	8 BAM fileerr:	194147_1.t	oam (errl	94147_1.bam);			🛃 File
	Run Info	Run Info			Dowr	nload		
	Alignment In	fo			4	bowtie2_report.txt		

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

×	💮 InSyBio Suite - DNA-Seq F	InSyBio Beta User 🔹 🗭 💡				
	InSyBio Interact	Job Status Job ID Submission Date Execution Time Input Data and Parameters				
	InSyBio ncRNASeq	CompleteD 5 Dec 18, 2019 7:57:02 AM 00 hours, 30 minutes, 02 seconds				
8	InSyBio Bionets	Variant Annotation Reports Trimmed FASTQ Files Trimmed FastQC Reports Bowtie2 Files Variant Calling Variant Annotation Ne	ext Actions			
	InSyBio Biomarkers	Variant Call Files Download	Download			
	InSyBio DNA-Seq	DNASeq Job-5 Variant Annotation file (hbr_1.vcf);				
	InSvBio Pipelines	DNASeq Job-5 Variant Annotation file (hbr_2.vcf);	🛓 File			
	InSyBio DataStore	DNASeq Job-5 Variant Annotation file (hbr_3.vcf);				

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

× 💮 InSyBio Suite - DNA-Seq F	ipeline Results	🚍 🙆 🌲 🛛 InSyBio Beta User 🔹 💭 🍞
inSyBio Interact	Job Status         Job ID         Submission Date         Execution Time         Input Data and Parameters           < Dashboard         commutered         5         Dec 18, 2019 7:57:02 AM         00 hours, 30 minutes, 02 seconds         1	
InSyBio Bionets	Variant Annotation Reports Trimmed FASTQ Files Trimmed FastQC Reports Bowtie2 Files Variant Calling Var	iant Annotation Next Actions
InSyBio Biomarkers	Missense Variant Vep Files	Download
InSvBio DNA-Seg	DNASeq Job-5 Filtered missense_variants Variant Annotation file (hbr_1_missense_annotations.vcf);	🛓 File
InSyBio Pipelines	DNASeq Job-5 Filtered missense_variants Variant Annotation file (hbr_2_missense_annotations.vcf);	🛓 File
InSyBio DataStore	DNASeq Job-5 Filtered missense_variants Variant Annotation file (hbr_3_missense_annotations.vcf);	🛓 File
	Protein Altering Variants	Download
	DNASeq Job-5 Filtered protein_altering_variants and AF < 0.05 Variant Annotation file (hbr_1_filtered_annotations.vcf);	File
	DNASeq Job-5 Filtered protein_altering_variants and AF < 0.05 Variant Annotation file (hbr_2_filtered_annotations.vcf);	E File
	DNASeq Job-5 Filtered protein_altering_variants and AF < 0.05 Variant Annotation file (hbr_3_filtered_annotations.vcf);	E File
	All Variants Downld	ad
	DNASeq Job-5 Variant Annotation file (hbr_1_annotations.vcf);	File
	DNASeq Job-5 Variant Annotation file (hbr_2_annotations.vcf);	File

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.

×	💮 InSyBio Suite - DNA-Seq F		InSyBio Beta User	•		
	InSyBio Interact	Job Status Job ID Submission Date Execution Time Input Data and Parameters				
	InSyBio ncRNASeq	CompleteD 5 Dec 18, 2019 7:57:02 AM 00 hours, 30 minutes, 02 seconds				
8	InSyBio <b>Bionets</b>	Variant Annotation Reports Trimmed FASTQ Files Trimmed FastQC Reports Bowtie2 Files Variant Calling Variant Annotation	N	lext Actions		
0	InSyBio <b>Biomarkers</b>	Continue your Analysis in InSyBio Suite				
	InSyBio DNA-Seq	Significant Gene Files Download Next Action				
		Significant Gene File - Condition 1 💽 File				
	InSyBio Pipelines					
	InSyBio DataStore					

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..

# How to get InSyBio DNASeq

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

To purchase InSyBio DNASeq commercial version 3.0 please contact us at <u>sales@insybio.com</u>.

### About Us

InSyBio Ltd is a bioinformatics pioneer company (<u>www.insybio.com</u>) 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 Ltd - Any InSyBio information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the InSyBio Ltd. A draft of the proposed document should accompany any such request. InSyBio Ltd reserves the right to deny approval of external usage for any reason.

Copyright 2022 InSyBio Ltd. Reproduction without written permission is completely forbidden.