



InSyBio

Intelligent Systems Biology

User Manual

www.insybio.com

Analyze coding & non-coding RNAs with InSyBio ncRNASeq

July 2024

Insybio Suite v3.3

Introduction

ncRNASeq is an RNA analysis tool for the prediction and analysis of:

- Coding RNAs
- non-coding RNAs
- miRNA target genes
- Bulk RNA-sequencing data
- single-cell RNA-sequencing data

Non-coding RNA genes are RNA sequences transcribed from DNA, but not translated to proteins. Their identification as well as the identification of the genes they regulate is a promising research area.

InSyBio ncRNASeq enables users to analyze non-coding RNAs. Users can search and analyze the RNA sequence of their interest. They can also analyze a full sequences dataset derived from online available databases, experimental sequencing techniques or computational in silico techniques.

With InSyBio ncRNASeq you can predict and analyze RNA genes and miRNA target genes by combining a variety of sequential, structural and functional information, and using a high-performance machine-learning technique. The RNA analysis is conducted by the calculation of the 58 most informative features described in the literature, and the miRNA-miRNA targets analysis is conducted by the calculation of the 124 most informative ones. InSyBio ncRNASeq also provides results storage in its knowledge base, equipped with information retrieval tools, to allow users to produce and extract their datasets.

With InSyBio ncRNASeq you can:

- a) Calculate 58 RNA genes-related features
- b) Predict miRNAs
- c) Calculate 124 miRNA target site features
- d) Predict miRNA target sites
- e) Search stem-loop and mature miRNAs

- f) Search transcripts and genes
- g) Search transcripts and genes for potential miRNA targets
- h) Predict miRNA targets
- i) Apply our processing pipeline to your RNASeq data and perform Differential Expression Analysis
- j) Identify different types of novel small non-coding RNAs (e.g. snoRNAs, miRNAs, tRNA fragments etc) from your raw RNA-sequencing data
- k) Apply our processing pipeline to your single-cell RNASeq data and perform Differential Expression Analysis, cell clustering and additional analyses (eg. cell-cell communication, identification of cell differentiation patterns, deconvolution).

non -coding RNA Analytics

ncRNA Feature Calculation

You can calculate 58 informative features for non-coding RNAs by supplying their sequence in fasta format. These features include sequential, thermodynamical and structural properties of the RNA sequences.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	35	ncrna15_12_	3/16/22 3:22 PM	3/16/22 3:22 PM	3/16/22 3:22 PM	View Results
Completed	34	ncrna15_12_	12/15/21 10:00 AM	12/15/21 10:00 AM	12/15/21 10:00 AM	View Results
Completed	33	ncrna 15_12	12/15/21 9:48 AM	12/15/21 9:48 AM	12/15/21 9:48 AM	View Results
Completed	32	ncrna14_12	12/14/21 10:01 AM	12/14/21 10:01 AM	12/14/21 10:01 AM	View Results
Completed	31	test	6/4/21 8:11 AM	6/4/21 8:11 AM	6/4/21 8:11 AM	View Results
Completed	29	75 sequences including pre-miRNAs, random cds and snoRNAs	3/4/21 4:43 PM	3/4/21 4:43 PM	3/4/21 4:43 PM	View Results
Completed	28	75 sequences including pre-miRNAs, random cds and snoRNAs	3/1/21 10:17 PM	3/1/21 10:17 PM	3/1/21 10:17 PM	View Results
Completed	27	75 sequences including pre-miRNAs, random cds and snoRNAs	1/4/21 6:06 PM	1/4/21 6:06 PM	1/4/21 6:06 PM	View Results

To start the calculation:

Select from the menu “Insybio ncRNASeq” → “non-coding RNA Analytics” → “ncRNA Feature Calculation”:

- Upload a new file of sequences in fasta format. You are redirected to the Data Store where step-by-step instructions guide you, or
- Select a file from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.

Batch calculations of many sequences are allowed. Just put the sequences in one file in fasta format.

Status	Process ID	Information	Submission Date	Start Execution
Completed	11	test	11/30/18 9:51 AM	11/30/18 9:51 AM
Completed	9	test	11/15/18 8:59 PM	11/15/18 8:59 PM
Completed	8	sequences75_premiRNAs_cds_snoRNAs2222	11/8/18 2:35 PM	11/8/18 2:35 PM
Completed	7	75 sequences including pre-miRNAs, random cds and snoRNAs	11/8/18 8:48 AM	11/8/18 8:49 AM
Completed	6	test	11/7/18 12:04 PM	11/7/18 12:04 PM
Pending	3	75 sequences including pre-miRNAs, random cds and snoRNAs	11/11/19 11:01 AM	-

To view the results:

By starting a calculation the ncRNA Feature Calculation dashboard is updated with the submitted job, there you can view the status of your current and previous ncRNA feature calculations. You can select the View Details at the Actions column and view the calculated features at completion of the calculation.

InSyBio Suite Beta - ncRNA Feature Calculation Results

Job Status: **COMPLETED** | Job ID: 1 | Submission Date: Aug 17, 2018 7:04:12 AM | Execution Time: 00 hours, 02 minutes, 35 seconds

Export Results

Sequence	G+C	AU	AA	AC	AG	AU	CA	CC	CG	CU
> hsa-mir-26a-1 MI000083 GUGGCCUGUCAAGUAUCCAGGAUAGGCGUGGACGGUCCAAUGGGCCUUAUUCUUGGUAUACUUGCACGGGGACGC	55.844	44.156	3.947	3.947	5.263	5.263	6.579	6.579	3.947	6.579
> random_sequence_from_cds_1 GAGGGCAGGGGGCAGUCCAAUCCAGGCUUGUAGUCUGUCCAGGGGUGGGUGCCGCCCGGACAGGGCAGACUGUCCUUGUGGCGGUGCA	69.072	30.928	1.042	4.167	8.333	0	10.417	9.375	4.167	6.25
> snoRNA_1 AAAGUGAGUGAUGAAUAGUUCUGGGCAUUGAAUUAUUUUUUGAUUAAACCCUAAACUCUGAAGUCC	32.857	67.143	14.493	2.899	5.797	11.594	2.899	4.348	0	5.797
> hsa-mir-32 MI000090 GGAGAUUUGCACAUUACUAGUUGCAUUGUUGCCAGGCUCAUUGCAUUAUUGUGUGUGAUUUUUC	38.571	61.429	4.348	4.348	4.348	11.594	8.696	1.449	1.449	2.899
> hsa-mir-199a-1 MI000242 GCCAACCCAGUGUUCAGACUACCGUUCAGGAGGUCUCUCAAUGUGUACAGUUGUCGACAUUGGUUAGGC	50.784	49.296	2.857	7.143	10	2.857	11.429	5.714	0	7.143
> hsa-mir-148a MI000253 GAGGCAAGUUCUGAGACACUCGACUCUGAGUAGUAGAAGUCAGUCCACUACAGAACUUUGUUCUC	45.588	54.412	5.97	8.955	11.94	2.985	7.463	1.493	1.493	10.448

Showing 1 entries

The results are presented on your screen in a browse-able table or you can download them as a TAB delimited txt file.

For each non-coding RNA, its sequence and its 58 features are presented.

The description of the supported features for the characterization of the non-coding RNAs is the following:

Feature	ABBR
2 Aggregate Dinucleotide Frequencies (%G+C ratio, %A+U ratio)	G + C, A + U
16 dinucleotide frequencies (%XY) such that X,Y \in Σ [A,C,G,U]	AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, UU
MFE Index 1 = $dG/\%(C+G)$	MFE1
MFE Index 2 = $dG/\text{number_of_stems}$, where each stem is at least 3 continuous base pairs in the structure	MFE2
MFE Index 3 = $dG/\text{number_of_loops}$, where number_of_loops is the number of the loops in the secondary structure	MFE3
MFE Index 4 = $dG/\text{total_bases}$	MFE4
MFE Index 5 = $dG/\%(A+U)$ ratio	MFE5
Adjusted Minimum Free Energy of folding $dG = \text{MFE}/L$, where MFE is the minimum free energy of the structure as calculated by the Vienna fold routine	dG
Adjusted base pairing propensity $dP = \text{total_bases}/L$, where L is the length of the structure and total_bases the number of base pairs in the structure	dP
Adjusted base pair distance dD	dD
Adjusted shannon entropy dQ	dQ
Positional Entropy dPs: a new introduced attribute which estimates the structural volatility of the secondary structure	PosEntropy
Normalized Ensemble Free Energy	EAFE
Structural Diversity	Div/ty
Frequency of MFE structure	Freq

Feature	ABBR
Diff = $ MFE-EFE /L$ where, EFE is the ensemble free energy	Diff
Structure Enthalpy dH	dH
Normalized Structure Enthalpy dH/L	dH/L
Structure Entropy dS	dS
Normalized Structure Entropy dS/L	dS/L
Melting Temperature Tm	Tm
Normalized Structure Enthalpy TH/L	Tm/L
X-Y is the number of (X-Y) base pairs in the secondary structure	A-U /L, G-C /L, G-U /L
Average base pair per stem	Avg_BP_stems
%(A-U)/n_stems, %(G-C)/n_stems, %(G-U)/n_stems.	(A-U)/n_stems, (G-C)/n_stems, (G-U)/n_stems
Ratio G/C ,where G,C is the number of G,C bases	G/C
BP is the total number of base pairs and GC,GU,AU the number of respective base pairs	BP/GC, BP/GU, BP/AU
Length of the sequence	Len
Centroid Energy: RNA folding related attribute calculated by the Vienna RNA package	DE/L
Centroid Distance: RNA folding related attribute calculated by the Vienna RNA package	CE_dist
5 statistical features	zG, zP, zD, zQ, zSP
Topological descriptor dF	dF

miRNA Prediction

You can predict pre-miRNAs and discriminate them between pseudo-hairpins and other molecules providing RNA sequences in fasta format. The prediction of pre-miRNAs and pseudo-hairpins is accomplished through the application of a novel methodology which combines Genetic Algorithms with epsilon-SVR techniques. Genetic Algorithms were used to optimize the feature subset which should be used as inputs and the parameters C, sigma and epsilon of epsilon SVR models. The accuracy of this technique in predicting pre-miRNAs is 95%. A sequence is predicted as other if the minimum free energy is more than -15 kcal/mol or the number of base pairs is less than 18.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	36	ncrna15_12_	3/16/22 3:26 PM	3/16/22 3:26 PM	3/16/22 3:26 PM	View Results
Completed	30	75 sequences including pre-miRNAs, random cds and snoRNAs	3/4/21 4:49 PM	3/4/21 4:50 PM	3/4/21 4:50 PM	View Results
Completed	14	75 sequences including pre-miRNAs, random cds and snoRNAs	11/11/19 11:36 AM	11/11/19 11:36 AM	11/11/19 11:36 AM	View Results
Completed	12	sequences10_premiRNAs_cds_snoRNAs	11/30/18 9:51 AM	11/30/18 9:51 AM	11/30/18 9:51 AM	View Results
Completed	10	test	11/15/18 9:00 PM	11/15/18 9:00 PM	11/15/18 9:00 PM	View Results
Completed	5	sequences75_premiRNAs_cds_snoRNAs2222	9/27/18 7:41 AM	9/27/18 7:41 AM	9/27/18 7:41 AM	View Results
Completed	4	75 sequences including pre-miRNAs, random cds and snoRNAs	9/26/18 11:18 AM	9/26/18 11:18 AM	9/26/18 11:18 AM	View Results
Completed	2	75 sequences including pre-miRNAs, random cds and snoRNAs	8/17/18 7:11 AM	8/17/18 7:11 AM	8/17/18 7:11 AM	View Results

To start the calculation:

Select from the menu “Insybio ncRNASeq” → “non-coding RNA Analytics” → “miRNA Prediction”:

- Upload a new file of sequences in fasta format. You are redirected to the Data Store where step-by-step instructions guide you.
- Select a file from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.

Batch calculations of many sequences are allowed. Just put the sequences in one file in fasta format.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	12	sequences10_premiRNAs_cds_snoRNAs	11/30/18 9:51 AM	11/30/18 9:51 AM	11/30/18 9:51 AM	View Results
Completed	10	test	11/15/18 9:00 PM	11/15/18 9:00 PM	11/15/18 9:00 PM	View Results
Completed	5	sequences75_premiRNAs_cds_snoRNAs2222	9/27/18 7:41 AM	9/27/18 7:41 AM	9/27/18 7:41 AM	View Results
Completed	4	75 sequences including pre-miRNAs, random cds and snoRNAs	9/26/18 11:18 AM	9/26/18 11:18 AM	9/26/18 11:18 AM	View Results
Completed	2	75 sequences including pre-miRNAs, random cds and snoRNAs	8/17/18 7:11 AM	8/17/18 7:11 AM	8/17/18 7:11 AM	View Results
Pending	14	75 sequences including pre-miRNAs, random cds and snoRNAs	11/11/19 11:36 AM	-	-	View Details

To view the results:

By starting a calculation the “miRNA Prediction” dashboard is updated with the status of the new job, there you can view the status of your current and previous miRNA prediction. At completion of the prediction, you can select the View Results in the Actions column and view the calculated features.

Sequence	Prediction Score	Prediction	G+C	A+U	AA	AC
> hsa-mir-26a-1 MI000083 GUGGCCUCGUUCAAGUUAUCCAGGUAAGGUGUGCAGGUCCAAUUGGGCCUUAUUUUGGUUAUCUUGCACGGGGACGC	1.02896	miRNA	55.844	44.156	3.947	3.947
> random_sequence_from_cds_1 GAGGGCAGGGGGCACAGUCCAAUCUCCAGGCUUUGUAGUCUCCAGGGGUGGGUGCCGCCGCGGAGCGGCAGACAGUGUCUGUGUGGCCUGGCACA	-0.893914	pseudomiRNA	69.072	30.928	1.042	4.167
> snoRNA_1 AAAGUGAGUGAUAAUAGUUCUGUGGCAUUAUAAUUAUUUUUAUUAUAAACCCUAAACUCUGAAGUCC	NaN	other	32.857	67.143	14.493	2.899
> hsa-mir-32 MI000098 GGAGAUAUUGCACAUACUAAGUUGCAUUGUUGCAGGCGCCUCAUUGCAUUUAGUGUGUGUGUAUUUUC	1.06856	miRNA	38.571	61.429	4.348	4.348
> hsa-mir-199a-1 MI000242 GCCAACCCAGUGUUCAGACUACUGUUCAGGAGGUCUCAUUGUGUACAGUAGUCUGCACAUGGUGUAGGC	0.92389	miRNA	50.784	49.296	2.857	7.143
> hsa-mir-148a MI000253 GAGGCAAGUUCUGAGACACUCCGACUCUGAGUAGUAGAAGUCAGUCACUACAGAAUUUGUCUC	1.17143	miRNA	45.588	54.412	5.97	8.955

The results are presented on your screen in a browseable table or you can download them as a TAB delimited txt file.

For each non-coding RNA, its sequence, its calculated confidence score, the prediction of whether it is a miRNA, a pseudo-hairpin or other and its 58 features are presented.

miRNA Target site Feature Calculation

You can calculate 124 features for every pair of a miRNA and its potential target site within an mRNA. These features include sequential, thermodynamical and structural properties of the miRNA:mRNA pair.

The screenshot displays the InSyBio Interact web interface. The sidebar on the left contains the following navigation items:

- InSyBio Interact
- InSyBio ncRNASeq
 - non-coding RNA Analytics
 - Prediction of ncRNAs and miRNA targets.
 - ncRNA Feature Calculation** (Feature calculation module for 58 miRNA genes-related features.)
 - miRNA Prediction (Prediction module for pre-miRNAs.)
 - miRNA Target site Feature Calculation** (Feature calculation module for 124 miRNA target features.)
 - miRNA Target site Prediction (Prediction module for miRNA targets.)
 - miRNA Target Prediction (Prediction module for miRNA targets.)
 - ncRNASeq Knowledge Base (miRNA and transcript search.)
 - RNA-Seq Data Analysis (Preprocessing and differential expression analysis of FASTQ files.)
 - Single Cell RNA-Seq Data Analysis

The main panel shows the 'miRNA Target site Feature Calculation' workflow. It includes input fields for mRNA Target Sequences (mimas462) and miRNA Sequences (mimas462), and buttons for 'Select file from Data Store' and 'Go to Data Store to Upload File'. A 'Start calculation' button is visible at the bottom right.

Below the input fields is a table of completed processes:

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	16	mRNAs: mimas462, miRNAs: mimas462	3/16/22 3:28 PM	3/16/22 3:28 PM	3/16/22 3:32 PM	View Results
Completed	14	mRNAs: mimas462, miRNAs: mimas462	3/4/21 5:22 PM	3/4/21 5:22 PM	3/4/21 5:44 PM	View Results
Completed	13	mRNAs: mimas462, miRNAs: mimas462	11/11/19 11:51 AM	11/11/19 11:51 AM	11/11/19 12:36 PM	View Results
Completed	11	mRNAs: test, miRNAs: test	11/30/18 9:52 AM	11/30/18 9:52 AM	11/30/18 10:23 AM	View Results
Completed	9	mRNAs: targetsbsa-miR-374-5pTCL1B-001.fa, miRNAs:	11/15/18 9:01	11/15/18 9:01	11/15/18 9:02	View Results

To start the calculation:

Select from the menu “InSyBio ncRNASeq” → “non-coding RNA Analytics” → “miRNA Target Features Calculation” and then:

- Upload a new file of mRNA binding sites sequences and a new file of miRNA sequences, both in fasta format. The mRNA target site of the first file and every miRNA of the second file are considered as a miRNA:mRNA pair. You are redirected to the Data Store where step-by-step instructions guide you for both files uploading.
- Or Select a file of mRNA binding sites sequences and a file of miRNA sequences, both in fasta format from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.

Batch feature calculation of many miRNA:mRNA pairs with a single run is allowed. Just put the mRNA binding sites sequences in the first file and miRNA sequences in the second file in fasta format.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	11	mRNAs: test, miRNAs: test	11/30/18 9:52 AM	11/30/18 9:52 AM	11/30/18 10:23 AM	View Results
Completed	9	mRNAs: targetshsa-miR-324-5pTCL1B-001.fa, miRNAs: miRNAs:hsa-miR-324-5pTCL1B-001.fa	11/15/18 9:01 PM	11/15/18 9:01 PM	11/15/18 9:02 PM	View Results
Completed	8	mRNAs: mrnas462, miRNAs: mirnas462	11/8/18 1:45 PM	11/8/18 1:45 PM	11/8/18 5:51 PM	View Results
Completed	3	mRNAs: mrnas462, miRNAs: mirnas462	9/26/18 11:21 AM	9/26/18 11:21 AM	9/26/18 12:00 PM	View Results
Completed	1	mRNAs: genes_5_5_0_shuffled_targets, miRNAs: genes_5_5_0_miRNAs	8/17/18 7:13 AM	8/17/18 7:13 AM	8/17/18 7:33 AM	View Results
Pending	13	mRNAs: mirnas462, miRNAs: mrnas462	11/11/19 11:51 AM	-	-	View Details

To view the results:

By starting a new calculation the “miRNA Target Site Feature Calculation” dashboard is updated with the new job, there you can view the status of your current and previous miRNA Target Features Calculations. At completion of the calculation, you can select the View Results in the Actions column and view the calculated features.

Job Status	Job ID	Submission Date	Execution Time	Input Data and Parameters	Export Results
COMPLETED	3	Sep 26, 2018 11:21:20 AM	00 hours, 39 minutes, 03 seconds		Export Results

miRNA Sequence	Target Sequence	mats	matos	mat	gcmats	gcmatos	gcmat	aumats	aumatos	aumat	unps	unpos	unp	gus	guos	gu	miss	m
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUAAUCUGAA	> NM_004456EZH220478051 Homo sapiens TGAATTTGCAAAGTACTGTA	9	2	11	3	1	4	6	1	7	-2	22	20	0	0	0	-2	
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUAAUCUGAA	> NM_004456EZH220478051 Homo sapiens TTCAGGAACCTCGACTACTGTG	8	6	14	3	3	6	5	3	8	0	16	16	2	2	4	-2	
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUAAUCUGAA	> NM_181833NF217220301 Homo sapiens TACAGAGATTCTCTGCCTCA	4	3	7	2	2	4	2	1	3	8	22	30	0	0	0	8	
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUAAUCUGAA	> NM_001039111TRIM7117890240 Homo sapiens ACAACATTGCTTAAGTCTACCTCA	1	5	6	0	2	2	1	3	4	14	21	35	0	2	2	14	
> [hsa-miR-101] Homo cantane	> NM_001039111TRIM7117890240 Homo cantane	9	3	12	3	2	5	6	1	7	-2	25	23	0	0	0	-2	

First Previous 1 2 3 4 5 ... 8538 Next Last Show 25 entries Showing 1 to 25 of 213,444 entries

The results are presented on your screen in a browse-able table or you can download them as a TAB delimited txt file.

For each miRNA:mRNA pair, the miRNA sequence, the mRNA binding site sequence and the 124 miRNA::mRNA pair features are presented.

The description of the supported features for the characterization of the miRNA::mRNA pair is the following:

Feature	ABBR	Category
number of matches in seed part	mats	structural
number of matches in out-seed part	matos	structural
total number of matches	mat	structural
number of GC matches in seed part	gcmats	structural
number of GC matches in out-seed part	gcmatos	structural
total number of GC matches	gcmat	structural
number of AU matches in seed part	aumats	structural
number of AU matches in out-seed part	aumatos	structural
total number of AU matches	aumat	structural
number of mismatches in seed part	unps	structural
number of mismatches in out-seed part	unpos	structural
total number of mismatches	unp	structural
number of GU wobble pairs in seed part	gus	structural
number of GU wobble pairs in out-seed part	guos	structural
total number of GU wobble pairs	gu	structural
number of other mismatches in seed part	miss	structural
number of other mismatches in out-seed part	misos	structural
total number of other mismatches	mis	structural
number of bulges in seed part	buls	structural

Feature	ABBR	Category
number of bulges in out-seed part	bulos	structural
total number of bulges	bul	structural
number of loops in seed part	symls	structural
number of loops in out-seed part	symlos	structural
total number of loops	syml	structural
number of asymmetric loops in seed part	asymls	structural
number of asymmetric loops in out-seed part	asymlos	structural
total number of asymmetric loops	asyml	structural
length of largest bulge	maxbul	structural
number of bulges of length 1-7 and greater than 7 in seed part (8 features)	cbul1s, cbul2s, cbul3s, cbul4s, cbul5s, cbul6s, cbul7s, cbul8s	structural
number of bulges of length 1-7 and greater than 7 in out-seed part (8 features)	cbul1os, cbul2os, cbul3os, cbul4os, cbul5os, cbul6os, cbul7os, cbul8os	structural
number of symmetric loops of length 1-7 and greater than 7 in seed part (8 features)	csl1s, csl2s, csl3s, csl4s, csl5s, csl6s, csl7s, csl8s	structural
number of symmetric loops of length 1-7 and greater than 7 in out-seed part (8 features)	csl1os, csl2os, csl3os, csl4os, csl5os, csl6os, csl7os, csl8os	structural
number of asymmetric loops of length 1-7 and greater than 7 in seed part (8 features)	cas1s, cas2s, cas3s, cas4s, cas5s, cas6s, cas7s, cas8s	structural
number of asymmetric loops of length 1-7 and greater than 7 in out-seed part (8 features)	cas1os, cas2os, cas3os, cas4os, cas5os, cas6os, cas7os, cas8os	structural
proportion of A, C, G, U in the target sequence (4	aper, cper, gper,	structural

features)	upper	
distance from the start of the seed part to the last match of the out-seed part	dist	structural
seed score obtained by the sum of pair scores in the seed region. GC and AU with 5, GU with 2 and the others with -3	scores	structural
out-seed score obtained by the sum of pair scores in the out-seed region. GC and AU with 5, GU with 2 and the others with -3	scoreos	structural
free energy of the seed part	mfes	thermodyna mic
free energy of the out-seed part	mfeos	thermodyna mic
free energy of the total miRNA-mRNA alignment structure	mfe	thermodyna mic
free energy of the target sequence	mfet	thermodyna mic
normalized free energy of the target sequence= $(-1 * \text{free energy of the target sequence}) / \log(\text{length of target} * \text{length of miRNA})$	nmfe	thermodyna mic
difference in the free energies of the total miRNA-perfect target alignment structure and the total miRNA-mRNA alignment structure	dmfe	thermodyna mic
positions from 1 to 20 with a GC match, an AU match, a GU match or a mismatch (20 features)	pos1, pos2, pos3, pos4, pos5, pos6, pos7, pos8, pos9, pos10, pos11, pos12, pos13, pos14, pos15, pos16, pos17, pos18, pos19, pos20	positional
terminal (position 8) base match	match8	positional
positional pair score obtained by the sum of the product of the weight and the corresponding pair score throughout the total miRNA-mRNA alignment structure. G:C and A:U are awarded with 5, G:U with 1, all other mismatches with -3 and the mismatches containing gaps with -1. Positional weight is 1 for all non-seed positions and 2 for all seed positions.	s106	positional

Feature	ABBR	Category
matrix score obtained by the sum of the diagonal elements in the matrix formed by the miRNA and its target. WC pairs: 5, Wobble pairs: 2, Inserts: -1, Deletes: -1, Symmetric mismatches: -3, Mismatches: -2	score	positional
deviation of the positional pair score with the score obtained with a perfect target	ds108	positional
deviation of the matrix score with the score obtained with a perfect target	ds109	positional
existence of the 10 most frequent nucleotide sequence 'words' with lengths 4, 5, 6, 7, 8 from the seed sequence of the miRNAs of our dataset	ugag, cagu, agug, agguag, aggua, aggu, gguag, ggua, guag, ugcu	'motif'

miRNA Target site Prediction

You can computationally validate miRNA targets. The computational intelligent technique, which was applied for the prediction of miRNAs (hybrid combination of Genetic Algorithms and epsilon-SVRs), and 124 informative features are used.

The screenshot displays the InSyBio ncRNASeq web interface. The sidebar on the left contains the following navigation options:

- InSyBio Interact
- InSyBio ncRNASeq
 - non-coding RNA Analytics
 - Prediction of ncRNAs and miRNA targets.
 - ncRNA Feature Calculation
 - Feature calculation module for 58 miRNA genes-related features.
 - miRNA Prediction
 - Prediction module for pre-miRNAs.
 - miRNA Target site Feature Calculation
 - Feature calculation module for 124 miRNA target features.
 - miRNA Target site Prediction
 - Prediction module for miRNA targets.
 - miRNA Target Prediction
 - Prediction module for miRNA targets.
 - ncRNASeq Knowledge Base
 - miRNA and transcript search.
 - RNA-Seq Data Analysis
 - Preprocessing and differential expression analysis of FASTQ files.
 - Single Cell RNA-Seq Data Analysis

The main form shows the following fields and options:

- miRNA Target Sequences:
- Filename:
- Buttons: [Select file from Data Store](#), [Go to Data Store to Upload File](#)
- miRNA Sequences:
- Filename:
- Buttons: [Select file from Data Store](#), [Go to Data Store to Upload File](#)
- [Start calculation](#)

The results table below shows the execution status of various jobs:

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	17	mRNAs: mrnas462, miRNAs: mirnas462	3/16/22 4:18 PM	3/16/22 4:19 PM	3/16/22 4:22 PM	View Results
Completed	15	mRNAs: mrnas462, miRNAs: mirnas462	3/4/21 5:24 PM	3/4/21 5:24 PM	3/4/21 5:47 PM	View Results
Completed	12	mRNAs: , miRNAs: test	11/30/18 9:54 AM	11/30/18 9:54 AM	11/30/18 9:56 AM	View Results
Completed	10	mRNAs: test, miRNAs: test	11/15/18 9:29 PM	11/15/18 9:29 PM	11/16/18 12:08 AM	View Results
Error	7	mRNAs: , miRNAs: pseudomi1848	9/27/18 9:36 AM	9/27/18 9:36 AM	11/30/18 10:11 AM	View Details

To start the prediction:

Select from the menu “InSyBio ncRNASeq” → “non-coding RNA Analytics” → “miRNA Target Site Prediction” and then:

- Upload a new file of candidate mRNA target binding sites sequences and a new file of miRNA sequences, both in fasta format. The mRNA target site of the first file and every miRNA of the second file are considered as a miRNA:mRNA pair. You are redirected to the Data Store where step-by-step instructions guide you for both files uploading.
- Or Select a file of candidate mRNA target binding sites sequences and a file of miRNA sequences, both in fasta format from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.

Batch predictions of many miRNA:mRNA pairs with a single run are allowed. Just put the candidate mRNA target binding sites sequences in the first file and miRNA sequences in the second file in fasta format.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	11	mRNAs: test, miRNAs: test	11/30/18 9:52 AM	11/30/18 9:52 AM	11/30/18 10:23 AM	View Results
Completed	9	mRNAs: targetshsa-miR-324-5pTCL1B-001.fa, miRNAs: miRNAshsa-miR-324-5pTCL1B-001.fa	11/15/18 9:01 PM	11/15/18 9:01 PM	11/15/18 9:02 PM	View Results
Completed	8	mRNAs: mrnas462, miRNAs: mirnas462	11/8/18 1:45 PM	11/8/18 1:45 PM	11/8/18 5:51 PM	View Results
Completed	3	mRNAs: mrnas462, miRNAs: mirnas462	9/26/18 11:21 AM	9/26/18 11:21 AM	9/26/18 12:00 PM	View Results
Completed	1	mRNAs: genes_5_5_0_shuffled_targets, miRNAs: genes_5_5_0_miRNAs	8/17/18 7:13 AM	8/17/18 7:13 AM	8/17/18 7:33 AM	View Results
Pending	13	mRNAs: mirnas462, miRNAs: mrnas462	11/11/19 11:51 AM	-	-	View Details

To view the results:

By starting a calculation the “miRNA Target Site Prediction” dashboard is updated with the new job, where you can view the status of your current and previous miRNA Target Site Prediction. At completion of the calculation, you can select the View Results in the Actions column and view the predictions and calculated features.

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

Dashboard **COMPLETED** 4 Sep 26, 2018 11:29:30 AM 01 hours, 10 minutes, 43 seconds [Export Results](#)

miRNA Sequence	Target Sequence	Prediction Score	Prediction	mats	matos	mat	gcmats	gcmatos	gcmat	aumats	aumatos	aumat	unps	unpos	unp
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUACUGAA	> NM_004456EZ220478051 Homo sapiens TGAATTTGCAAAGTACTGTA	0.963256	Target	9	2	11	3	1	4	6	1	7	-2	22	28
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUACUGAA	> NM_004456EZ220478051 Homo sapiens TTCAGGAACCTCGAGTACTGTG	1.2725	Target	8	6	14	3	3	6	5	3	8	0	16	16
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUACUGAA	> NM_181833NF217220301 Homo sapiens TACAAGAGATTCTCCTGCCTCA	-0.786746	no Target	4	3	7	2	2	4	2	1	3	8	22	30
> [hsa-miR-101] Homo sapiens UACAGUACUGUGAUACUGAA	> NM_001039111TRIM7117890240 Homo sapiens ACAAATTGCTTAAGTCTACCTCA	-0.880751	no Target	1	5	6	0	2	2	1	3	4	14	21	35

First Previous 1 2 3 4 5 ... 8538 Next Last Show 25 entries Showing 1 to 25 of 213,444 entries

The results are presented on your screen in a browseable table or you can download them as a TAB delimited txt file.

For each miRNA:mRNA pair, the miRNA sequence, the mRNA binding site sequence, whether the miRNA:mRNA pairs share a targeting relation or not, the confidence score of the prediction and all 124 miRNA::mRNA are presented.

miRNA Target Prediction

You can computationally predict potential miRNA targets at given Genes or Transcripts and given miRNAs. BLAST is performed to find potential target sites, and then the computational intelligent technique, which was applied for the prediction of miRNAs (hybrid combination of Genetic Algorithms and epsilon-SVRs), and 124 informative features are used to calculate a prediction score.

The screenshot displays the InSyBio Suite - miRNA Target Prediction Tool interface. On the left is a navigation sidebar with the following sections:

- InSyBio Interact**
- InSyBio ncRNASeq**
 - non-coding RNA Analytics
 - Prediction of ncRNAs and miRNA targets.
 - ncRNA Feature Calculation (Feature calculation module for 58 miRNA genes-related features.)
 - miRNA Prediction (Prediction module for pre-miRNAs.)
 - miRNA Target site Feature Calculation (Feature calculation module for 124 miRNA target features.)
 - miRNA Target site Prediction (Prediction module for miRNA targets.)
 - miRNA Target Prediction (Prediction module for miRNA targets.)
 - ncRNASeq Knowledge Base (miRNA and transcript search.)
 - RNA-Seq Data Analysis (Preprocessing and differential expression analysis of FASTQ files.)
 - Single Cell RNA-Seq Data Analysis (Preprocessing and differential expression analysis of FASTQ files.)

The main panel contains search fields:

- Search miRNA:** A dropdown menu with "Select miRNA" and an "Add to list" button. Below it, a text area labeled "miRNAs List:" contains "hsa-miR-205-3p,hsa-miR-205-5p".
- Search Genes:** A search box with "GNAQ" entered and an "Add to list" button. Below it, a dropdown menu shows suggestions: "GNAQ", "GNAQ", and "GNAQP1". A text area labeled "Genes List:" is empty.

A "Queue new Process" button is located below the search fields.

At the bottom of the interface is a table showing the results of completed processes:

Status	Process ID	Information
Completed	163	miRNAs: hsa-miR-205-3p,hsa-miR-205-5p targets: GNAQ
Completed	162	miRNAs: hsa-let-7a-5p,hsa-let-7a-3p,hsa-let-7a-2-3p,hsa-let-7b-5p,hsa-let-7b-3p,hsa-let-7c-5p,hsa-let-7d-5p,hsa-let-7d-3p,hsa-let-7e-3p,hsa-let-7e-5p,hsa-let-7f-5p,hsa-let-7f-3p,hsa-let-7g-5p,hsa-let-7g-3p,hsa-let-7f-2-3p,hsa-let-7e-1-5p,hsa-let-7e-1-3p,hsa-let-7e-2-5p,hsa-let-7e-2-3p,hsa-let-7f-2-5p,hsa-let-7f-2-3p,hsa-let-7g-2-5p,hsa-let-7g-2-3p,hsa-let-7g-3-5p,hsa-let-7g-3-3p,hsa-let-7g-4-5p,hsa-let-7g-4-3p,hsa-let-7g-5-5p,hsa-let-7g-5-3p,hsa-let-7g-6-5p,hsa-let-7g-6-3p,hsa-let-7g-7-5p,hsa-let-7g-7-3p,hsa-let-7g-8-5p,hsa-let-7g-8-3p,hsa-let-7g-9-5p,hsa-let-7g-9-3p,hsa-let-7g-10-5p,hsa-let-7g-10-3p,hsa-let-7g-11-5p,hsa-let-7g-11-3p,hsa-let-7g-12-5p,hsa-let-7g-12-3p,hsa-let-7g-13-5p,hsa-let-7g-13-3p,hsa-let-7g-14-5p,hsa-let-7g-14-3p,hsa-let-7g-15-5p,hsa-let-7g-15-3p,hsa-let-7g-16-5p,hsa-let-7g-16-3p,hsa-let-7g-17-5p,hsa-let-7g-17-3p,hsa-let-7g-18-5p,hsa-let-7g-18-3p,hsa-let-7g-19-5p,hsa-let-7g-19-3p,hsa-let-7g-20-5p,hsa-let-7g-20-3p,hsa-let-7g-21-5p,hsa-let-7g-21-3p,hsa-let-7g-22-5p,hsa-let-7g-22-3p,hsa-let-7g-23-5p,hsa-let-7g-23-3p,hsa-let-7g-24-5p,hsa-let-7g-24-3p,hsa-let-7g-25-5p,hsa-let-7g-25-3p,hsa-let-7g-26-5p,hsa-let-7g-26-3p,hsa-let-7g-27-5p,hsa-let-7g-27-3p,hsa-let-7g-28-5p,hsa-let-7g-28-3p,hsa-let-7g-29-5p,hsa-let-7g-29-3p,hsa-let-7g-30-5p,hsa-let-7g-30-3p,hsa-let-7g-31-5p,hsa-let-7g-31-3p,hsa-let-7g-32-5p,hsa-let-7g-32-3p,hsa-let-7g-33-5p,hsa-let-7g-33-3p,hsa-let-7g-34-5p,hsa-let-7g-34-3p,hsa-let-7g-35-5p,hsa-let-7g-35-3p,hsa-let-7g-36-5p,hsa-let-7g-36-3p,hsa-let-7g-37-5p,hsa-let-7g-37-3p,hsa-let-7g-38-5p,hsa-let-7g-38-3p,hsa-let-7g-39-5p,hsa-let-7g-39-3p,hsa-let-7g-40-5p,hsa-let-7g-40-3p,hsa-let-7g-41-5p,hsa-let-7g-41-3p,hsa-let-7g-42-5p,hsa-let-7g-42-3p,hsa-let-7g-43-5p,hsa-let-7g-43-3p,hsa-let-7g-44-5p,hsa-let-7g-44-3p,hsa-let-7g-45-5p,hsa-let-7g-45-3p,hsa-let-7g-46-5p,hsa-let-7g-46-3p,hsa-let-7g-47-5p,hsa-let-7g-47-3p,hsa-let-7g-48-5p,hsa-let-7g-48-3p,hsa-let-7g-49-5p,hsa-let-7g-49-3p,hsa-let-7g-50-5p,hsa-let-7g-50-3p,hsa-let-7g-51-5p,hsa-let-7g-51-3p,hsa-let-7g-52-5p,hsa-let-7g-52-3p,hsa-let-7g-53-5p,hsa-let-7g-53-3p,hsa-let-7g-54-5p,hsa-let-7g-54-3p,hsa-let-7g-55-5p,hsa-let-7g-55-3p,hsa-let-7g-56-5p,hsa-let-7g-56-3p,hsa-let-7g-57-5p,hsa-let-7g-57-3p,hsa-let-7g-58-5p,hsa-let-7g-58-3p,hsa-let-7g-59-5p,hsa-let-7g-59-3p,hsa-let-7g-60-5p,hsa-let-7g-60-3p,hsa-let-7g-61-5p,hsa-let-7g-61-3p,hsa-let-7g-62-5p,hsa-let-7g-62-3p,hsa-let-7g-63-5p,hsa-let-7g-63-3p,hsa-let-7g-64-5p,hsa-let-7g-64-3p,hsa-let-7g-65-5p,hsa-let-7g-65-3p,hsa-let-7g-66-5p,hsa-let-7g-66-3p,hsa-let-7g-67-5p,hsa-let-7g-67-3p,hsa-let-7g-68-5p,hsa-let-7g-68-3p,hsa-let-7g-69-5p,hsa-let-7g-69-3p,hsa-let-7g-70-5p,hsa-let-7g-70-3p,hsa-let-7g-71-5p,hsa-let-7g-71-3p,hsa-let-7g-72-5p,hsa-let-7g-72-3p,hsa-let-7g-73-5p,hsa-let-7g-73-3p,hsa-let-7g-74-5p,hsa-let-7g-74-3p,hsa-let-7g-75-5p,hsa-let-7g-75-3p,hsa-let-7g-76-5p,hsa-let-7g-76-3p,hsa-let-7g-77-5p,hsa-let-7g-77-3p,hsa-let-7g-78-5p,hsa-let-7g-78-3p,hsa-let-7g-79-5p,hsa-let-7g-79-3p,hsa-let-7g-80-5p,hsa-let-7g-80-3p,hsa-let-7g-81-5p,hsa-let-7g-81-3p,hsa-let-7g-82-5p,hsa-let-7g-82-3p,hsa-let-7g-83-5p,hsa-let-7g-83-3p,hsa-let-7g-84-5p,hsa-let-7g-84-3p,hsa-let-7g-85-5p,hsa-let-7g-85-3p,hsa-let-7g-86-5p,hsa-let-7g-86-3p,hsa-let-7g-87-5p,hsa-let-7g-87-3p,hsa-let-7g-88-5p,hsa-let-7g-88-3p,hsa-let-7g-89-5p,hsa-let-7g-89-3p,hsa-let-7g-90-5p,hsa-let-7g-90-3p,hsa-let-7g-91-5p,hsa-let-7g-91-3p,hsa-let-7g-92-5p,hsa-let-7g-92-3p,hsa-let-7g-93-5p,hsa-let-7g-93-3p,hsa-let-7g-94-5p,hsa-let-7g-94-3p,hsa-let-7g-95-5p,hsa-let-7g-95-3p,hsa-let-7g-96-5p,hsa-let-7g-96-3p,hsa-let-7g-97-5p,hsa-let-7g-97-3p,hsa-let-7g-98-5p,hsa-let-7g-98-3p,hsa-let-7g-99-5p,hsa-let-7g-99-3p,hsa-let-7g-100-5p,hsa-let-7g-100-3p,hsa-let-7g-101-5p,hsa-let-7g-101-3p,hsa-let-7g-102-5p,hsa-let-7g-102-3p,hsa-let-7g-103-5p,hsa-let-7g-103-3p,hsa-let-7g-104-5p,hsa-let-7g-104-3p,hsa-let-7g-105-5p,hsa-let-7g-105-3p,hsa-let-7g-106-5p,hsa-let-7g-106-3p,hsa-let-7g-107-5p,hsa-let-7g-107-3p,hsa-let-7g-108-5p,hsa-let-7g-108-3p,hsa-let-7g-109-5p,hsa-let-7g-109-3p,hsa-let-7g-110-5p,hsa-let-7g-110-3p,hsa-let-7g-111-5p,hsa-let-7g-111-3p,hsa-let-7g-112-5p,hsa-let-7g-112-3p,hsa-let-7g-113-5p,hsa-let-7g-113-3p,hsa-let-7g-114-5p,hsa-let-7g-114-3p,hsa-let-7g-115-5p,hsa-let-7g-115-3p,hsa-let-7g-116-5p,hsa-let-7g-116-3p,hsa-let-7g-117-5p,hsa-let-7g-117-3p,hsa-let-7g-118-5p,hsa-let-7g-118-3p,hsa-let-7g-119-5p,hsa-let-7g-119-3p,hsa-let-7g-120-5p,hsa-let-7g-120-3p,hsa-let-7g-121-5p,hsa-let-7g-121-3p,hsa-let-7g-122-5p,hsa-let-7g-122-3p,hsa-let-7g-123-5p,hsa-let-7g-123-3p,hsa-let-7g-124-5p,hsa-let-7g-124-3p

To start the prediction:

Select from the menu “InSyBio ncRNASeq” → “non-coding RNA Analytics” → “miRNA Target Prediction” field and then:

- Select the miRNAs and the Genes you want to search for potential targets by searching in our Database and adding them to the miRNA List and Genes List or add them manually to their Lists and separating them with commas.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	89	miRNAs: hsa-miR-6126 targets: ZIK1	11/11/19 3:02 PM	11/11/19 3:02 PM	11/11/19 3:02 PM	View Results
Completed	88	miRNAs: mmu-miR-3072-3p,mmu-miR-7051-3p,mmu-miR-3968,mmu-miR-8106,mmu-miR-99a-3p,mmu-miR-21a-3p,mmu-miR-3110-5p,mmu-miR-505-3p,mmu-miR-7091-5p,mmu-miR-337-5p,mmu-miR-18a-3p,mmu-miR-1949,mm... targets: ZIK1	2/11/19 12:11 PM	6/6/19 11:21 AM	6/6/19 3:39 PM	View Results
Completed	87	miRNAs: hsa-miR-576-3p, hsa-miR-140-5p, hsa-miR-522-5p, hsa-miR-1298-5p, hsa-miR-133a-3p, hsa-miR-4743-3p, hsa-miR-557, hsa-miR-548ao-3p, hsa-miR-5088-5p, hsa-miR-4649-5p, hsa-miR-665, hsa-miR-3622b-... targets: NELL2, SERPINI1, SMOCI, FGF2, MMRN2, PRSS3, VEGFB, ADAM21, ADAMTSL4, C1QTNF4, CCL3L3, COL4A2, LAMB1	11/29/18 3:40 PM	11/29/18 3:40 PM	11/29/18 3:52 PM	View Results
Completed	86	miRNAs: hsa-miR-6126, hsa-miR-1200, hsa-let-7a-2-3p, hsa-miR-106b-3p targets: ZIK1, A1BG-AS1, FGGY	11/29/18 3:39 PM	11/29/18 3:39 PM	11/29/18 3:39 PM	View Results
Completed	85	miRNAs: hsa-miR-6126 targets: ZIK1	11/29/18 3:09 PM	11/29/18 3:09 PM	11/29/18 3:09 PM	View Results
Error	84	miRNAs: targets: ZIK1	11/29/18 3:08 PM	11/29/18 3:08 PM	11/29/18 3:08 PM	View Details

To view the results:

By starting a calculation the “miRNA target Prediction” dashboard is updated with the new job’s information, you can view the status of your current and previous miRNA Target Predictions. After the calculation, you can select the View Results in the Actions column and view the results.

Mirna Target Prediction Tool Results InSyBio Beta User

Job Status	Job ID	Submission Date	Execution Time	Input Data and Parameters	Actions
COMPLETED	89	Nov 11, 2019 3:02:12 PM	00 hours, 00 minutes, 02 seconds	Results Download all target sites found Download miRNA-target genes scores	

miRNA	Gene	Score	Actions
hsa-miR-6126	ZIK1	1.169	Details

miRNA	Gene	Transcript	Score	Actions
hsa-miR-6126	ZIK1	ZIK1-002	0.817	Utr Sequence
hsa-miR-6126	ZIK1	ZIK1-001	0.817	Utr Sequence
hsa-miR-6126	ZIK1	ZIK1-004	1.517	Utr Sequence
hsa-miR-6126	ZIK1	ZIK1-003	1.527	Utr Sequence

The results are presented on your screen in a browseable table, with each miRNA and gene pair in a row with their confidence score. By pressing Details at the Actions Column the specific scores between the miRNA and the gene’s transcripts can be

viewed. If no target sites are found “No targets found!” is presented at the score column. If one or more target sites are found you can view its UTR sequence, with the target sites of the miRNA highlighted. Multiple target sites are marked with green color and unique target sites are marked with light blue.

- Gene show page InSyBio Beta User

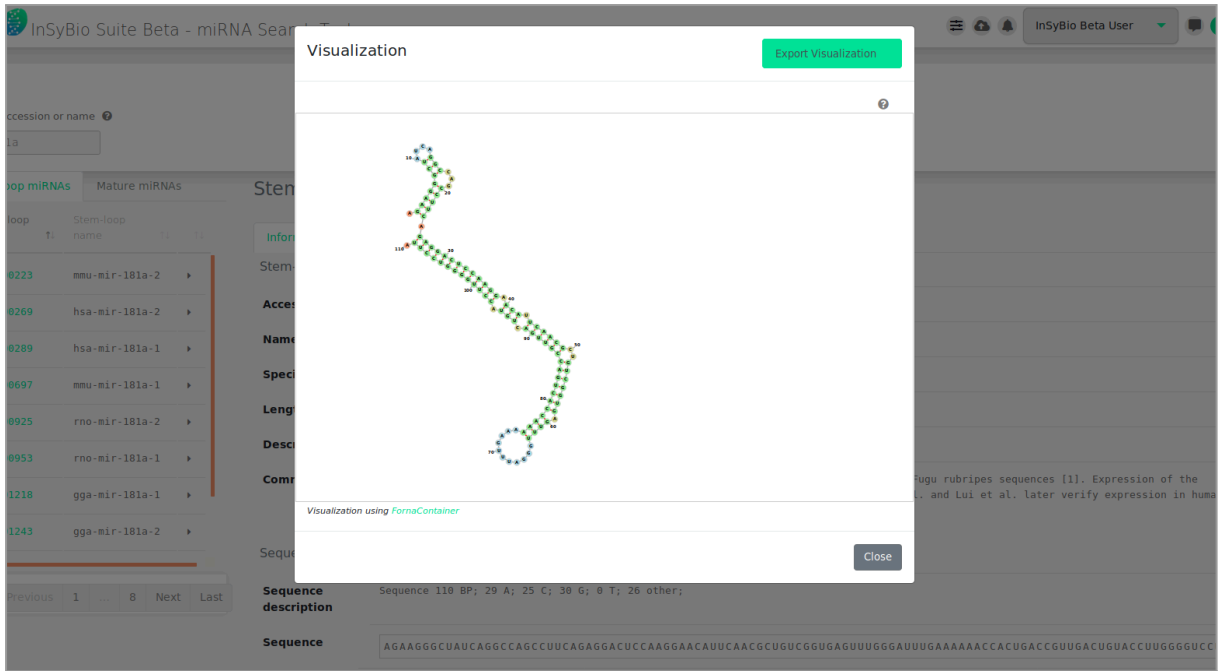
← Mirna Target Prediction Tool Results

miRNA	hsa-miR-6126
Gene	ZIK1
Transcript	ZIK1-001
miRNA-Gene Score	1.169
miRNA-Transcript Score	0.817
Number of target sites	11
3'UTR sequence	<pre> 1 AGGCCTCATGAATGCAGCAATGTGGAAGCGCCTCAACTCAAGATCTATCATCATTAGCTCCTGAAAGCCACACTTA 80 81 AGTAGAGCCTTAGACCTACAGGAAAGTGCTGTCTCTGTAGTATTGTAGCAGTAGAGAGCCTTTGTGAGGGAGCCATCTG 160 161 CCTGAAGTTGAACCTCATTCTTCTCTGTTCTCTGGTAGAAACCATCTACCCCTACCACTTGCACAGTGGCACTGGT 240 241 CACTCCTATGTGCTAAGACAAGGCAGACATCTGTGTCTCTTAAGTCTTTGGAGGAAATCTTGAGCAGTCTAAGCCCT 320 321 TAGAGAAAATTCATTCTTTTTCTGACTGATCACAGCATACGTGTGACCCAGTTTGGGTGAGGAGGCCAGCCCTTGGTT 400 401 CTGCTGGACACTTATGTGCAAGGATCCCTTCATGTAAATTCITGGTCTCACAGACACTTGGTCATCTCTCCAGCCTCC 480 481 ATGTCACCACGTGGTGAATGGCTGCCTCACATTGCTCCAGTTGTGCACATAAAAAGCCTTATATTGAATCACCTGT 560 561 AGCTCTGGGGTCTGTTACTGTGTGGGGTGGCTGGGAGACAGACTCAACTCTATATGAAGGAATGGATGGCTTTTGTG 640 641 GGCCCTGCAGGAAAGTAAGATGACAGAGTAATTCTAATTCTGGTTTGGTCATACCTGCTTGGCTACCTAAAATCTCC 720 721 AGGAAAAATGCAAGGTTTGGTTATTCTAATTGTGGCCGGATCCCTATTCTTTCTGTGAGACTAGAGGTCATCTCGA 800 801 GGAGAGGCCAGCTGTTATGACAAGCATGTGTCTCAGGGAATAGGACAATTTATTCCATTGTTCCAGAGGATGTCAT 880 881 ATGATGCCAGTGTCTGAGAAGCTTTTATGGGGTCTATAAAGAGGCATGCCCTGATCAAAACATCCATAGGCCG 960 961 ATGTCACGCAGAAACACCGGAGTCCATGTGAAGTGAATTTGGTACAGAAATACCTGGGTATTCTGTACTGTGTGTA 1040 1041 CTGTAGCAAACTAGTTGGAATGTGCCTCTATAAAAAGTACATTTACAAATCTCCCGTGAAGTGGCTTTGAGCAGTCAT 1120 1121 AAGGACCTAGAATCTGTGTATGTCCAATAGCTGAGGTTATTTTACAGCAAAAATAATTAAGGGTTTTATTTTTTAATCT 1200 1201 TGTGGTTTTCTAGGTTGTTCACCTCAAGTGCATTGCTGTAGAGGCAGAAAAAGGAGGATAAAGATAACAGAAAGTCCAT 1280 1281 AAGGCCAGGGATGATTGATAGCTCTTGTGATTTCCACCAGTGTGCTGTTGCTCAAATTGCACAGCCTTCTATTGCTG 1360 1361 CCAACATTTCTGCATGGAGGACTCATGGTGGCCCTCCCCAGGCCCTGAAGAGAGAGTGCAGTCAACATGAGATTGCTA 1440 1441 GGCATTCTGGTTCTGAAAGTGGGTGATCAGATACTTTATTGTGAAACATGTTTACAACTCTTCTGATGTGAAAGT 1520 1521 ACATGCCATAGTTTACATCCATTTATGGTGTATAATTTGAAGAGTTTGTATACAAGCCTGTGAAACATAATCATGATC 1600 1601 ATGAACATATTCATGATTCACCTCTTGGTGTTTTACAATCTGCGGTGACTTCCAGGCCCTCAGGAGTCCCTGCTCATT 1680 1681 ACTTCCCTACAGGAGAATAGTTTGTGTTTCTAGGATTTTATGTGAATTGAACGTAATAACTTACTCCATTTTCTC 1760 </pre>

Score : 1.7294313303229796
TTCCCTCATGTAATTTCTTGGTCT-CACAT
||||| ||| ||||| |||
---AGAGG-----CGGCCCGGAAGUG---

Score : 1.5224538611539185
TGACACTTGTCTTCTCCAGCCTCATG
||||| ||||| |||
-----AGAGGCCCGCGGAAGUG

You can download all target sites found as a txt file.



Mature miRNAs and references

miRNA accession or name Show results

Stem-loop miRNAs Mature miRNAs Stem-loop: MI0000269 hsa-mir-181a-2

Stem-loop id	Stem-loop name	Information	Mature miRNAs	References			
MI0000223	mmu-mir-181a-2	Accession	Name	Sequence	FASTA	Evidence	Experiment
MI0000269	hsa-mir-181a-2	MIMAT0000256	hsa-mir-181a-5p	39 aacaucaacgcugucggugagu 61	Download	Experimental	cloned [2,4-6]
MI0000289	hsa-mir-181a-1	MIMAT0004558	hsa-mir-181a-2-3p	77 accacugaccguagacuacc 98	Download	Experimental	cloned [4]

For the mature miRNAs related to the stem-loop of interest you can view their accession, name and sequence. Concerning the sequence, you can download the fasta format. You can also view the evidence of each mature miRNA, which can be experimental, or by the similarity of the related stem-loop to another stem-loop or found in the literature.

miRNA accession or name Show results

Stem-loop miRNAs Mature miRNAs **Stem-loop: MI0000269 hsa-mir-181a-2**

Stem-loop id	Stem-loop name	TI	TI
MI0000223	mmu-mir-181a-2		
MI0000269	hsa-mir-181a-2		
MI0000289	hsa-mir-181a-1		
MI0000697	mmu-mir-181a-1		
MI0000925	rno-mir-181a-2		
MI0000953	rno-mir-181a-1		
MI0001218	gga-mir-181a-1		
MI0001243	gga-mir-181a-2		

Information Mature miRNAs **References**

Links to external database entries

Database	External Link
	MI0000269
	mir-181
	MIR181A2
	MIR181A2

Publications

- Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP: **Vertebrate microRNA genes**; Science. 299:1540(2003). [PubMed]
- Dostie J, Mourelatos Z, Yang M, Sharma A, Dreyfuss G: **Numerous microRNPs in neuronal cells containing novel microRNAs**; RNA. 9:180-186(2003). [PubMed]
- Weber MJ: **New human and mouse microRNA genes found by homology search**; FEBS J. 272:59-73(2005). [PubMed]
- Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Jj, Sander C, Zavolan M, Tuschl T: **A mammalian microRNA expression atlas based on small RNA library sequencing**; Cell. 129:1401-1414(2007). [PubMed]
- Lui WO, Pourmand N, Patterson BK, Fire A: **Patterns of known and novel small RNAs in human cervical cancer**; Cancer Res. 67:6031-6043(2007). [PubMed]
- Marton S, Garcia MR, Robello C, Persson H, Trajtenberg F, Pritsch O, Rovira C, Naya H, Dighiero G, Cayota A: **Small RNAs analysis in CLL reveals a deregulation of miRNA**

You can also view references for the miRNA of interest. There are external links to other databases (MIRBASE, ENTEZGENE, HGNC, RFAM, MGI, and WORMABASE) and publications.

Mature miRNA information

miRNA accession or name Show results

Stem-loop miRNAs **Mature miRNAs** **Mature: MIMAT0000210 mmu-miR-181a-5p**

Mature id	Mature name	TI	TI
MIMAT0000210	mmu-miR-181a-5p		
MIMAT0000210	mmu-miR-181a-5p		
MIMAT0000256	hsa-miR-181a-5p		
MIMAT0000256	hsa-miR-181a-5p		
MIMAT0000270	hsa-miR-181a-3p		
MIMAT0000660	mmu-miR-181a-1-3p		

Information Stem-loop miRNAs References

Accession MIMAT0000210

Name mmu-miR-181a-5p

Sequence 14 aacauucaacgcugucggugagu 36

FASTA Download

Evidence Experimental

Experiment cloned [2,4], Illumina [5-6]

Similarity MI0000223

For the Mature miRNA you can view their accession, name and sequence. Concerning the sequence, you can download the fasta format. You can also view the

evidence of each mature miRNA, which can be experimental, or by similarity of the related stem-loop to another stem-loop or found in the literature.

For the stem-loop related to the mature mi-RNA of interest you can view its accession, name, species, length, description and comments. Concerning its sequence, you can download the fasta format and view the sequence description, the sequence and the secondary structure in dot-bracket notation. You can view the visualization of the secondary structure by clicking the “Visualization” button, this visualization is performed with FornaContainer. It is the Minimum Free Energy (MFE) structure.

You can also view references for the mature miRNA of interest. There are external links to other databases (MIRBASE, ENTEZGENE, HGNC, RFAM, MGI, and WORMABASE) and publications.

Transcript Search

You can search transcripts and genes by giving a transcript accession or name or part of them. Choosing the transcript or gene of those returned, its show page is shown.

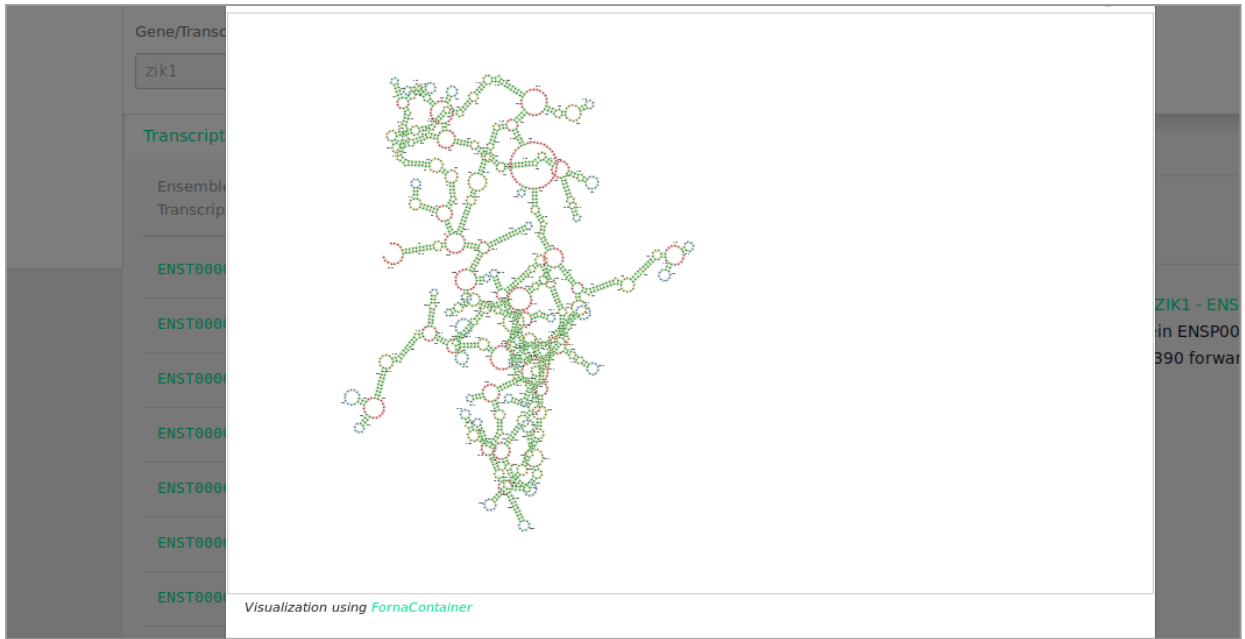
Transcripts information

The screenshot displays the InSyBio ncRNASeq interface. At the top, there is a search bar containing 'zik1' and a green 'Show results' button. Below the search bar, there are two tabs: 'Transcripts' (selected) and 'Genes'. A table lists several transcripts with columns for 'Ensemble Transcript id', 'Transcript name', and 'T1'. The transcript ZIK1-004 (ENST00000307468) is highlighted. To the right of the table, the detailed information for ZIK1-004 (ENST00000307468) is shown. This information includes:

- Name - Source:** ZIK1-004 (HGNC transcript name)
- Gene:** This transcript is a product of gene [ZIK1 - ENSG00000171649](#)
- Protein:** This transcript corresponds to protein [ENSP00000303820](#).
- Location:** Chromosome 19: 57584260-57592390 forward strand
- Transcription Start Site (TSS):** 57584260
- Length:** 2510
- Transcript Support Level (TSL):** TSL:1
- Gencode annotation:** GENCODE basic
- GC content:** 47.45 %
- Biotype:** protein_coding
- Status:** Known
- Annotation method:** Havana
- Version:** ENST00000307468.4
- Description:** zinc finger protein interacting with K protein 1 [Source:HGNC Symbol;Acc:HGNC:33104 [External Link to HGNC](#)]

 At the bottom of the transcript information, there is a '3'UTR Visualization' section with a green 'Visualization' button and a 'Download' button.

For the Transcript you can view its name-source, gene, protein, location, transcription start site (TSS), length, transcription support level (TSL), Gencode annotation, GC content, biotype, status, annotation method and version description. Concerning its 3'UTR sequence, you can download the fasta format and view the sequence description, the sequence and the secondary structure in dot-bracket notation. You can view the visualization of the secondary structure by clicking the "Visualization" button, this visualization of the secondary structure is performed with FornaContainer. It is the Minimum Free Energy (MFE) structure.



Ensemble Transcript id	Transcript name	Information
ENST00000307468	ZIK1-004	<p>3'UTR sequence</p> <p>GAGTGTACAGTCAAAGGCAGGTTTCACACAGAAGACTCAATCCTGTGAGATGTGTGTCCCAAGTCTGAAAGATATT TTGCATCTAGCTGATCTCCCTGGGCAGAAACCATACTTGGTTGGAGAATGTACAAACCATCACCAGCACCAGAAGCATCA CAGTGCAAGAAATCCTTGAAGAGGGACATGGACAGAGCCTCATATGTGAAGTGTGCTTATCTGTATGTCATTGGAAGC CTTTTCGCAAAATGGGAGGTTGGAAAAGGACCTCCAGCCATGTTGCGGCTTCTGAGGTCCCTGGTCTTTCTGGAGGCAAG AAACCCGGCACAATTACTGAATGTGGGGAGGACATTGCGAGTCAAAAAAGTCATTACAAGTCAAGTGAATGTGGGAAGGC TTCCAGGCACAAACACACTCCTGTTTACCATCCAAGAGTCTACACTGGAAAAAGCCTTATGAGTGTAGCAAAATGTGGGA AAGCCTTCCGTGGCAAGTACTCACTTGTTCAGCACAGAGAGTCCATACTGGAGAAAGGCCTTGGGAGTGAATGAATGT GGAATAATCTTATGCCAAACCTCCACCTGAATGATCATCGGAGAATCCACACGGAGAAAGGCCTTATGAGTGCAGCGA ATGTGGAAATATTTAGACAAAACCTCCAGCCTGTTGACCACCAAAAAATACACACTGGAGCAAGGCCTTATGAGTGA GCCAGTGTGGGAAATCCTTATGCCAAAAAGCCACCTTGTAAACACCAAAGAGTTACACTGGAGAAAGGCCTTATAAG TGTGTTGAATGTGGAAATCTTTAGTCAAAGTGCCATTCTTAATCAACACCGAAGAATTCACACTGGAGCAAAGCCTTA TGAGTGTGGCCAGTGTGGGAAATCCTTATGCAAAAAGCTACCTCATTAAACACCAAGAGATTCACACTGGAGAAAGGC CTTATAAGTGTGGTACTGTGGGAAATCCTTATGCAAAAAGCTCCTTATGCAAAAAGCTCCTTATGCAAAAAGCTCCTT AAGCCTTATGAGTGTGGCCAGTGTGGAAAGTCTTATGCAAAAAGCTCCTTATGCAAAAAGCTCCTTATGCAAAAAGCT AGAAAAGCCTTATGAGTGAACAAATGTGGGAAATCCTTATGCAAAAAGCTCCTTATGCAAAAAGCTCCTTATGCAAAA ACACATAGAGGCCTCATGAATGCAGCAAAATGTGGAAAGGCCTTCAACTCAAGATCTATCATCATTTAGCTCCTGAAAGTC CACACTAAGTAGAGCCTTAGACCTACAGGAAAGTGTGCTCTGTAGTATTGTAGCAGTAGAGAGCCTTTGTGAGGGA GCCATCTGCCTGAAGTTGAACCTCATTCTTCTTCTGTTCTCTGGTAGAAACCATCACCTCTACCACCTTGACAGTGG GCACTGGTCACTCCTATGTGCTAAGACAAGGCAGACATCTGTGTCTCTTAAGTCTTTGGAGGAAATCTTGGAGCATC TAAGCCTTATGAGAAATTCATTCTTTTTCTGACTGATCACAGCATACGTTGTGACCCAGTTTGGGTACAGGAGGCCAG CTTTGGTTCTGTGGACACTTATGTGCAAGGATCCCTTCATGTAATTTCTGGTCTCACATGACACTTGGTCACTTCTTC CAGCCTCATGTACACAGTGGTGAATGGCTGCTCACATTGCTCCAGTTTGTGCACTAATAAAAGCCTTATATTTGAAT CTACCTGTAGTCTTGGGTTCTGTTTACTGTGTGGGGTGGCTGGGAGACAGACTTCAACTCTATATGAAGGAATGGATGG CTTTGTGGCCTTCGAGGAAAGTAAAGTACAGAGTAATTCTAATTTCTGGTTTGGTCACTATGCTTTGCTACCTAA AATCTCTAGGAAAAATGCAAGGTTTGGTTATTCTAATTTGTGGCTGGATCCCTATTTCTTGTGAGACTAGAGGT CATCTGAGGAGAGGCCAGCTGTTATGACAAGCATGTGTGCTTCAGGGAATAGGACAATTTATCCATTGTTCCAGAG</p>
ENST00000456074	ZIK1P1-001	
ENST00000536878	ZIK1-002	
ENST00000597219	ZIK1-006	
ENST00000597850	ZIK1-001	
ENST00000598689	ZIK1-007	
ENST00000598726	ZIK1-008	
ENST00000599456	ZIK1-003	

First Previous 1 Next Last

Genes information

The screenshot shows the 'Gene Search Tool' interface. A search box contains 'zik1' and a 'Show results' button. The 'Genes' tab is active, displaying a table of search results:

Ensemble Gene id	Official Gene Symbol
ENSG00000171649	ZIK1
ENSG00000237426	ZIK1P1

The main panel shows details for 'Gene:ZIK1 ENSG00000171649' under the 'Information' tab:

- Name - Source:** ZIK1 (HGNC Symbol)
- Description:** zinc finger protein interacting with K protein 1 [Source:HGNC Symbol;Acc:HGNC:33104 External Link to HGNC]
- Location:** Chromosome 19: 57578456-57593777 forward strand
- Transcript count:** 8
- Biotype:** protein_coding
- Status:** Known
- Annotation method:** Annotation for this gene includes both automatic annotation from Ensembl and Havana External Link manual curation, see article External Link
- Version:** ENSG00000171649.11

A 'Transcript Table' tab is also visible, which is expanded in the second screenshot.

For the Genes you can view its name-source, description, location, transcript count, biotype, status, annotation method and version. Also, a Transcript Table is provided with the genes associated transcripts and links to their information.

This screenshot shows the same search results as the first image, but with the 'Transcript Table' tab selected. The table lists 8 transcripts associated with the ZIK1 gene:

#	Ensemble id	Name
1	ENST00000536878	ZIK1-002
2	ENST00000597219	ZIK1-006
3	ENST00000597850	ZIK1-001
4	ENST00000598689	ZIK1-007
5	ENST00000598726	ZIK1-008
6	ENST00000599456	ZIK1-003
7	ENST00000600053	ZIK1-005
8	ENST00000307468	ZIK1-004

RNA-Seq Data Analysis

Rna-Seq Differential Expression 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 HISAT2, and sorting with Samtools.
- C.** Quantification using FeatureCounts.
- D.** Differential Expression using Deseq2.

Firstly, the Pipeline uses Fastqc to create a report with the sequence quality, then trim the sequences accordingly using Trimmomatic and create new reports with Fastqc. Then using HISAT2 it creates the alignment SAM files, we sort them using SAMtools and transform them to BAM files. The BAM files are used as input for FeatureCounts, which creates text files with the quantity of each gene. In the end, DESeq2 performs Differential Expression Analysis for all the pairs of conditions using R.

We also offer a modification to the above pipeline, called ncRNA-Seq Differential Expression Pipeline, where the unaligned reads from the Alignment step are used to enhance the quantification files with known or predicted ncRNAs. This is done by finding all the contigs of the unaligned reads files using the AbySS Assembler, and then checking if these contigs are known ncRNAs (from a list of 6 ncRNA types: miRNA, pre-miRNA, tRNA, rRNA, snoRNA and tRf) or use our novel method of an EnsembleGASVR Classifier to predict if the contigs are possible ncRNAs. Then the quantity of the known and predicted ncRNAs is used to enhance the quantification files produced by featureCounts and continue the pipeline as described above.

To start the differential expression:

Click in the menu “InSyBio ncRNASeq” → “RNA-Seq Data Analysis” → “RNA-Seq Diff. Expression Pipeline Dashboard”, select the “Add new job” button and then:

- Select if you have Paired or Single Ended data.

The screenshot shows the InSyBio Suite - RNA-Seq Differential Expression Pipeline interface. The top navigation bar includes the InSyBio logo, the title "InSyBio Suite - RNA-Seq Differential Expression Pipeline", and the user "InSyBio Beta User". The left sidebar contains several menu items: "InSyBio Interact", "InSyBio ncRNASeq", "non-coding RNA Analytics", "ncRNASeq Knowledge Base", "RNA-Seq Data Analysis", "RNA-Seq Diff. Expression Pipeline Dashboard", "Single Cell RNA-Seq Data Analysis", "InSyBio Bionets", "InSyBio Biomarkers", "InSyBio DNA-Seq", "InSyBio Pipelines", and "InSyBio DataStore".

The main content area is titled "RNA-Seq Data:" and features a radio button selection for "Paired-end" (selected) and "Single-ended". Below this, there are three job configurations, each with a "Condition" field and a "Data:" section. The first configuration is for condition "hbr" and the second for "uhr". Each configuration includes fields for "Title Read 1", "Title Read 2", "Filename Read 1", and "Filename Read 2". There are also "Select from Data Store" and "Upload to Data Store" buttons for each read. A "Delete Pair" button is present for each pair, and an "Add Pair" button is at the bottom right of each configuration. At the bottom of the main area, there are "Options" for "Do you want to perform initial FastQC?" and "Do you want to perform trimming?".

The screenshot displays the InSyBio Suite interface for RNA-Seq Differential Expression Pipeline. The left sidebar contains navigation menus for 'InSyBio Interact', 'InSyBio ncRNASeq', 'non-coding RNA Analytics', 'ncRNASeq Knowledge Base', 'RNA-Seq Data Analysis', 'RNA-Seq Diff. Expression Pipeline Dashboard', 'Single Cell RNA-Seq Data Analysis', 'InSyBio Bionets', 'InSyBio Biomarkers', 'InSyBio DNA-Seq', 'InSyBio Pipelines', and 'InSyBio DataStore'. The main panel is titled 'RNA-Seq' and features radio buttons for 'Paired-end' and 'Single-ended'. Below this, the 'Data' section is divided into two conditions: 'hbr' and 'uhf'. Each condition includes a 'Title' and 'Filename' field, and buttons for 'Select from Data Store', 'Upload to Data Store', and 'Delete File'. There are also 'Add File' and 'Add Condition' buttons. At the bottom, the 'Options' section includes checkboxes for 'Do you want to perform initial FastQC?' and 'Do you want to perform trimming?', along with a dropdown menu for '--Select Action--'.

- Name Conditions/Group of files you want to compare.
- For each condition add single or paired files by:
 - Uploading a new file of Rna-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 Rna-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 on the initial Data.

Options

Do you want to perform initial FastQC

Do you want to perform trimming?

Alignment Options

Source for the reference genome *

Specify strand information:

- 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, either from our 4 built-in options (HumanGRCh37, HumanGRCh38, MouseGRCm38 and ZebrafishGRCz11), or
 - Upload new reference Genome files in fasta and gtf format. You are redirected to the Data Store where step-by-step instructions guide you for both files uploading.
 - Or Select two reference Genome files one in fasta and one in gtf format from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.

Alignment Options

Source for the reference genome *

Use a genome from Data Store

Select the reference genome (FASTA): *

Title: chr22 fasta

Filename: dsfile1573556494_9916.fa

Select from Data Store Upload to Data Store

Select the reference genome (GTF): *

Title: chr22 GTF

Filename: dsfile1573556655_8832.gtf

Select from Data Store Upload to Data Store

Alignment Options

Source for the reference genome *

Use a built-in genome

Select a reference genome: *

HumanGRCh38

Specify strand information:

Forward (FR)

- Select the strandness of your input files, Unstranded, Forward or Reverse.
- If more than 2 Conditions are selected, you can select which pairs of conditions to Differentially Express (all versus Control, all versus all or assign manually).

- Last but not least select either to perform the regular RNASeq Differential Expression Pipeline or the enhanced ncRNASeq Differential Expression Pipeline.

Which conditions do you want to compare? Set manually ▾

	Control ▾	Tumor ▾	-
	Control ▾	Treated ▾	-
Condition Pairs:	Tumor ▾	Treated ▾	-

+

RNASeq Analysis ncRNASeq 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 Rna-Seq Differential Expression Pipeline and view the Dashboard, where you can view the status of your current and previous Rna-Seq Differential Expression jobs.

Status	Job ID	Job Type	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step	Actions
Completed	62	ncRNASeq Analysis	hbr: 1. HBR rep1 read1, HBR rep1 read2, 2. HBR rep2 read1, HBR rep2 read2 uhr: 3. UHR rep1 read1, UHR rep1 read2, 4. UHR rep2 read1, UHR rep2 read2	3/14/22 9:59 AM	3/14/22 11:42 AM	Differential Expression Analysis	View Results	
Completed	61	RNASeq Analysis	Dox: 1. IonXpressRNA_007.Dox-1_small, 2. IonXpressRNA_013.Dox-2_small Lck: 3. IonXpressRNA_015.Lck-1_small, 4. IonXpressRNA_016.Lck-2_small Lyn: 5. IonXpressRNA_014.Lyn-1_small, 6. IonXpressRNA_012.Lyn-2_small	12/1/21 1:39 PM	12/1/21 1:39 PM	12/2/21 12:36 AM	Differential Expression Analysis	View Results
Error	60	RNASeq Analysis	Control: 1. 8212_3878 Howard 1, 2. 8212_30 Howard 2, 3. 8212_2430 Howard 3 Case: 4. 1009_062_3870 Howard 4, 5. 1009_062_2430 Howard 5, 6. GoHawks	11/3/21 11:40 AM	11/3/21 12:05 PM	11/5/21 3:20 AM	Differential Expression Analysis	View Details
Completed	58	RNASeq Analysis	Dox: 1. IonXpressRNA_001.Dox-1.fastq, 2. IonXpressRNA_002.Dox-2.fastq	10/27/21 10:48 AM	10/27/21 10:48 AM	10/28/21 5:12 AM	Differential Expression Analysis	View Results

After 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.

Job Status	Job ID	Submission Date	Execution Time	Input Data and Parameters
COMPLETED	1	May 6, 2019 7:55:09 AM	00 hours, 15 minutes, 49 seconds	

Deseq2 Reports

- Deseq2 Report File (.pdf) [Download]
- Job-1 DESeq2 pdf output [File]
- Deseq2 Report File (.png) [Download]
- HBR_UHRimages.zip [Image Folder]
- Deseq2 Report File (.csv) [Download]
- Job-1 DESeq2 output HBR_UHR_diffexpr-results-with-counts.csv (HBR_UHR_diffexpr-results-with-counts.csv); [File]
- Job-1 DESeq2 output HBR_UHR_diffexpr-results.csv (HBR_UHR_diffexpr-results.csv); [File]
- Job-1 DESeq2 output HBR_UHR_diffexpr-resultssignificant_pvalues.csv (HBR_UHR_diffexpr-results_significant_pvalues.csv); [File]

In Deseq2 reports tab you can download visual information and the Differential Expression calculated values for each pair compared.

FastQC Report	Download	View Html Page
Job-1 Fastqc zip file HBR rep1 read1	[Folder]	dsfile1557128487_9359_fastqc
Job-1 Fastqc zip file HBR rep1 read2	[Folder]	dsfile1557128516_9128_fastqc
Job-1 Fastqc zip file HBR rep2 read1	[Folder]	dsfile1557128550_6204_fastqc
Job-1 Fastqc zip file HBR rep2 read2	[Folder]	dsfile1557128587_1781_fastqc
Job-1 Fastqc zip file HBR rep3 read1	[Folder]	dsfile1557128617_6024_fastqc
Job-1 Fastqc zip file HBR rep3 read2	[Folder]	dsfile1557128647_9984_fastqc

In the Initial FastQC reports the FastQC reports of the input files can be downloaded.

Deseq2 Reports	Initial FastQC Reports	Trimmed FASTQ Files	Trimmed FastQC Reports	Alignment Files	Read Count Files	Next Actions
Trimmed FASTQ File					Download	
Job-1 trimmend paired file of HBR rep1 read1 (dsfile1557128487_9359_trimmed.gz);					File	
Job-1 trimmend paired file of HBR rep1 read2 (dsfile1557128516_9128_trimmed.gz);					File	
Job-1 trimmend paired file of HBR rep2 read1 (dsfile1557128550_6204_trimmed.gz);					File	
Job-1 trimmend paired file of HBR rep2 read2 (dsfile1557128587_1781_trimmed.gz);					File	
Job-1 trimmend paired file of HBR rep3 read1 (dsfile1557128617_6024_trimmed.gz);					File	
Job-1 trimmend paired file of HBR rep3 read2 (dsfile1557128647_9984_trimmed.gz);					File	
Job-1 trimmend paired file of UHR rep1 read1 (dsfile1557128760_6526_trimmed.gz);					File	

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

Deseq2 Reports	Initial FastQC Reports	Trimmed FASTQ Files	Trimmed FastQC Reports	Alignment Files	Read Count Files	Next Actions
Trimmed FastQC Report			Download		View Html Page	
s:51:"Job-1 after trimming Fastqc zip file HBR rep1 read1";			File	dsfile1557128487_9359_trimmed_fastqc		
s:51:"Job-1 after trimming Fastqc zip file HBR rep1 read2";			File	dsfile1557128516_9128_trimmed_fastqc		
s:51:"Job-1 after trimming Fastqc zip file HBR rep2 read1";			File	dsfile1557128550_6204_trimmed_fastqc		
s:51:"Job-1 after trimming Fastqc zip file HBR rep2 read2";			File	dsfile1557128587_1781_trimmed_fastqc		
s:51:"Job-1 after trimming Fastqc zip file HBR rep3 read1";			File	dsfile1557128617_6024_trimmed_fastqc		
s:51:"Job-1 after trimming Fastqc zip file HBR rep3 read2";			File	dsfile1557128647_9984_trimmed_fastqc		

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


The screenshot displays the 'Alignment Files' tab in a web interface. It is divided into three main sections: SAM File, BAM File, and Run Info. Each section has a 'Download' link. The SAM File section lists six files: HBR_1.sam, HBR_2.sam, HBR_3.sam, UHR_1.sam, UHR_2.sam, and UHR_3.sam. The BAM File section lists six files: HBR_1.bam, HBR_2.bam, HBR_3.bam, UHR_1.bam, UHR_2.bam, and UHR_3.bam. The Run Info section contains a single file named 'hisat2_report.txt'. Each file entry includes a green download icon and the text 'File'.


File Name	Download Link
Job-1 Hisat2 alignment file HBR_1.sam (HBR_1.sam);	File
Job-1 Hisat2 alignment file HBR_2.sam (HBR_2.sam);	File
Job-1 Hisat2 alignment file HBR_3.sam (HBR_3.sam);	File
Job-1 Hisat2 alignment file UHR_1.sam (UHR_1.sam);	File
Job-1 Hisat2 alignment file UHR_2.sam (UHR_2.sam);	File
Job-1 Hisat2 alignment file UHR_3.sam (UHR_3.sam);	File
BAM File	
Job-1 BAM fileHBR_1.bam (HBR_1.bam);	File
Job-1 BAM fileHBR_2.bam (HBR_2.bam);	File
Job-1 BAM fileHBR_3.bam (HBR_3.bam);	File
Job-1 BAM fileUHR_1.bam (UHR_1.bam);	File
Job-1 BAM fileUHR_2.bam (UHR_2.bam);	File
Job-1 BAM fileUHR_3.bam (UHR_3.bam);	File
Run Info	
Alignment Info	hisat2_report.txt

In the Alignment files tab, the HISAT2 alignment sam and bam files can be downloaded.

Read Count File	Download	Download Run Info File
Job-1 Feature counts file (HBR_1.counts);	 HBR_1.counts	 HBR_1.features.summary
Job-1 Feature counts file (HBR_2.counts);	 HBR_2.counts	 HBR_1.features.summary
Job-1 Feature counts file (HBR_3.counts);	 HBR_3.counts	 HBR_1.features.summary
Job-1 Feature counts file (UHR_1.counts);	 UHR_1.counts	 HBR_1.features.summary
Job-1 Feature counts file (UHR_2.counts);	 UHR_2.counts	 HBR_1.features.summary
Job-1 Feature counts file (UHR_3.counts);	 UHR_3.counts	 HBR_1.features.summary

In the Read Count Files tab the Count files for each sample can be downloaded.

Job Status	Job ID	Submission Date	Execution Time	Input Data and F
COMPLETED	79	Oct 2, 2019 8:56:41 AM	00 hours, 01 minutes, 56 seconds	

Predicted ncRNAs	Download
Predicted ncRNAs file	 File

If ncRNASeq Analysis is selected in the Predicted ncRNAs tab a tsv file with the found ncRNAs in the unaligned file is provided, with its name and predicted labels can be downloaded.

The screenshot displays the 'Next Actions' tab in the InSyBio Suite. The interface is organized into sections for downloading and acting on molecule quantification files. The top navigation bar includes tabs for 'Deseq2 Reports', 'Initial FastQC Reports', 'Trimmed FASTQ Files', 'Trimmed FastQC Reports', 'Alignment Files', 'Read Count Files', and 'Next Actions'. Below the navigation, the current analysis is identified as 'HBR_UHR'. The main content area is divided into two sections: 'Molecule Quantification Files per Condition' and 'Full Molecule Quantification File and Associated Labels'. Each section contains a list of files with a 'Download' button (a green arrow icon) and a 'Next Action' dropdown menu. The files listed are: 'Job-1 MQ file HBR_UHR_diffexpr-MQHBR.csv (HBR_UHR_diffexpr-MQHBR.csv);', 'Job-1 MQ file HBR_UHR_diffexpr-MQUHR.csv (HBR_UHR_diffexpr-MQUHR.csv);', 'Job-1 MQ file HBR_UHR_diffexpr-MQ.csv (HBR_UHR_diffexpr-MQ.csv);', and 'Job-1 label file HBR_UHR_diffexpr-labels.txt (HBR_UHR_diffexpr-labels.txt);'.

In the Next Action tab, Molecule Quantifications files, with the 10% most significant genes, for each comparison are provided. They can be downloaded or used as input in **InSyBio Bionets**, to construct gene correlation networks with the gene expressions of the genes found as statistically significantly differential expressed, and in **InSyBio Biomarkers**, to perform additional statistical analysis and build a classification model able to predict to which of the two conditions a potential new sample belongs.

single - cell RNA-Seq Data Analysis

single-cell RNA-Seq Differential Expression Pipeline

You can analyze single-cell RNA-Seq experiments. Alignment, read counts computation and additional secondary analysis are all performed in one job. Depending on the selected workflow, the single-cell RNA-Seq Differential Expression pipeline consists of the following 2 or 3 steps:

- Workflow 0 or 1:
 - Alignment and read counts computation using Cellranger count.
 - Further analysis using our single-cell Analysis.
- Workflow 2 or 3:
 - Alignment and read counts computation using Cellranger count pipeline for each different sample or different GEM well.
 - Aggregation of the Cellranger count runs using the Cellranger aggr pipeline.
 - Further analysis using our single-cell Analysis.

Firstly, the Pipeline uses the Cellranger count pipeline to perform the alignment and the read counts computation of the input fastq files. If the input fastq files are generated from different samples or different GEM wells, an extra step is performed. Specifically, the Cellranger aggr pipeline is used to aggregate the cellranger count runs for the creation of a single feature-barcode matrix and analysis. At the end, our single-cell Analysis script is used to perform additional secondary differential expression analysis.

To start the single-cell differential expression:

Click in the menu “InSyBio ncRNASeq” → “single-cell RNA-Seq Data Analysis” → “single-cell RNA-Seq Pipeline Dashboard”, select the “Add new job” button and then:

- Select your workflow.

☰ InSyBio Suite - single-cell RNA-Seq Differential Expression Pipeline
☰ 📁 🔔 InSyBio Beta User
🗨️ ?

Workflow One Sample, One GEM Well, Multiple Flowcells ⓘ

Input Data Files

Choose or upload to input your Fastq files to InSyBio single-cell RNA-Seq Differential Expression Pipeline tool following the rules:

- Fastq files must be in this name format: [Sample name]_S*_[Read Type]_001.fastq.gz (e.g. singlecell1_S1_R1_001.fastq.gz)
- Both R1 and R2 versions of each file must be present
- Fastq files of the same sample must have the same sample name
- Fastq files of different samples must have different sample name

Fastq File 1 ⓘ

Title1:

Filename 1:

📁 Select file from Data Store
➔ Go to Data Store to Upload File

Fastq File 2 ⓘ

Title2:

Filename 2:

📁 Select file from Data Store
➔ Go to Data Store to Upload File

Add File

Options

Transcriptome Human ⌵

Cluster annotation

Species: --Select Action-- ⌵

Tissue ⓘ --Select Action-- ⌵

- Upload your files of single-cell RNA-Seq Experiments in the following format:
 - Fastq files must be in this name format: [Sample name]_S*_[Read Type]_001.fastq.gz
 - Fastq files of the same sample must have the same sample name
 - Fastq files of different samples must have different sample name
- Select the transcriptome the input files belong to from our 3 built-in options (Human, Mouse, Human-mouse mixture).

- Select the species and tissue type of your sample for cluster annotation to be performed.
- Select if you want to manually configure the parameters of the pipeline. If you don't, our Default Options will be applied. Possible manual options are:
 - Expected number of recovered cells
 - BAM file generation
 - First filtering:
 - Minimum cells
 - Minimum features
 - Secondary filtering:
 - nFeature_RNA with lower and upper limits
 - nCount_RNA with lower and upper limits
 - Feature Extraction Method
 - Shared Nearest Neighbor (SNN) Graph
 - Clustering
 - Differentially expressed genes criteria
 - Plot for the top differentially expressed genes for each cluster
 - Genes for visualization

Advanced Options +

Expected number of recovered cells

BAM file generation



First filtering

Minimum cells:

Minimum features:

Secondary filtering

nFeature_RNA ? : Yes

Lower limit: 200

Upper limit: 10000

nCount_RNA ? : Yes

Lower limit:

Upper limit:

Feature Extraction Method: Umap

Shared Nearest Neighbor (SNN) Graph

k parameter (k-nearest-neighbor): 20

Clustering

Resolution parameter ? : 0.8


Differentially expressed genes criteria


Threshold (logfc):

Minimum Pct:

Plot for the top differentially expressed genes for each cluster

Number of top markers per cluster:

Average log2FC 

Genes for visualization All Custom 

- Submit your job pressing the respective button.

To view the results:

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

Status	Job ID	Job Type	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step	Actions
Error	1	RNASeq Analysis		2/9/22 1:10 PM	2/28/22 9:56 AM	2/27/22 7:28 PM	Single Cell Alignment	View Details
Completed	2	RNASeq Analysis		2/23/22 1:21 PM	2/28/22 6:51 AM	2/28/22 8:04 AM	Secondary Single Cell Analysis	View Results
Error	3	RNASeq Analysis			3/9/22 8:24 PM	2/28/22 5:58 PM	Single Cell Alignment	View Details
Running	4	RNASeq Analysis			3/15/22 10:08 AM	-	Secondary Single Cell Analysis	View Details

After the analysis, you can select the View Results at the Actions column and view the produced files, that are separated according to the step that they were produced.

InSyBio Suite - RNA-Seq Single Cell Pipeline Differential Expression Pipeline Results
InSyBio Beta User

	Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
Dashboard	COMPLETED	RNASeq Single Cell Analysis	2	Feb 23, 2022 1:21:53 PM	01 hours, 13 minutes, 17 seconds	i

Report | Summary | Additional Cell Statistics | Dot Plots Visualization | Feature Plots Visualization | Ridge Plots Visualization | Umap Plots Visualization | All Results Download

Single Cell Pipeline Report

Alignment of the sequencing reads in the provided FASTQ files to the selected reference transcriptome and read counts computation are performed with the Cellranger count pipeline. The outs folder contains the outputs of this step and includes the web_summary.html file which summarizes the results.

Secondary Single Cell Analysis

For the secondary single cell analysis quality control checks and filtering criteria are applied to the single cell data. With the Seurat Object the data are filtered using min.cells = 0 and min.features = 0.

min.cells: Include features detected in at least this many cells.
min.features: Include cells where at least this many features are detected.

An additional filtering step is performed with Seurat, keeping only cells that have unique feature counts and total number of molecules detected within a cell with the following limits:
nFeature_RNA = unique feature counts. lower limit: 100, upper limit: 3000
nCount_RNA = total number of molecules detected within a cell. lower limit: , upper limit:

The data are then normalized using the LogNormalize method, which normalizes the feature expression measurements for each cell by the total expression, multiplies this by a scale factor (10.000) and log-transforms the result.

2000 highly variable features that exhibit high cell-to-cell variation in our data are identified. Scaling is subsequently performed scaled, so that the mean expression across cells is 0 and variance across cells is 1. This last step is necessary for performing PCA on the data

The cells are clustered using a modularity optimization technique called Louvain algorithm with a resolution parameter of 1 (it sets the granularity of the downstream clustering) having firstly constructed the KNN graph (with k=30) based on the Euclidean distance in PCA space and using Jaccard similarity. Using the clustered data, non-linear dimensionality reduction is performed, producing the Umap plot.

In scRNA seq data analysis, differentially expressed features that define the clusters are called markers. To identify these markers, we firstly used the FindAllMarkers() function of the Seurat package, which identifies these markers for all clusters by comparing all clusters with each other. For this function we used parameters min.pct (a feature to be detected at a minimum percentage in either of the groups of cells) with value 0.1 and logfc.threshold (Limit testing to genes which show, on average, at least X-fold difference (log-scale) between the two groups of cells) with value 0.25. The matrices produced by these functions contain the genes as rows and these specific associated statistics for each gene as columns: P value, Average log2 Fold Change, Percentage of cells 1, Percentage of cells 2 and Adjusted P value.

The Dotplots include the differentially expressed genes that are only differentially expressed in one cluster of cells while sorting them by their p value.

The scCATCH package, a single cell Cluster-based annotation Toolkit for Cellular Heterogeneity is finally used to identify the cluster marker genes and creates the cluster annotations. We used the scCATCH() function which does the cluster annotation by matching the potential marker genes with known cell marker genes in a tissue-specific cell taxonomy reference database (CellMatch). We used the cancer type: and tissue types: Blood, Bladder.

The selected species was Human.

Results files description

Outs folder: The output files of the cellranger platform.

Web_summary.html: Variety of metrics such as Mean Reads per Cell, Median Genes per Cell, Valid Barcodes etc. At the analysis tab, t-SNE projection can be seen with UMI Counts or Clustered. Also, info about the Top features by cluster can be found.

Results of the secondary single cell analysis:

RidgePlots folder: Folder containing a Ridge Plot per gene you selected.

FeaturePlots folder: Folder containing a Feature Plot per gene you selected.

Dimensionality Reduction Plot folder:

Contains Umap.png: Umap projection plot of the clusters.

Markers folder:

markers_from_FindAllMarkers and markers_from_scCATCH: Markers(differentially expressed genes) and associated statistics (p-values, avg_log2FC etc) from FindAllMarkers and scCATCH functions respectively.

average_expression_of_genes.csv: Averaged expression values for every gene for every cluster.

Barcode-cluster.csv: Barcode-cluster matrix.

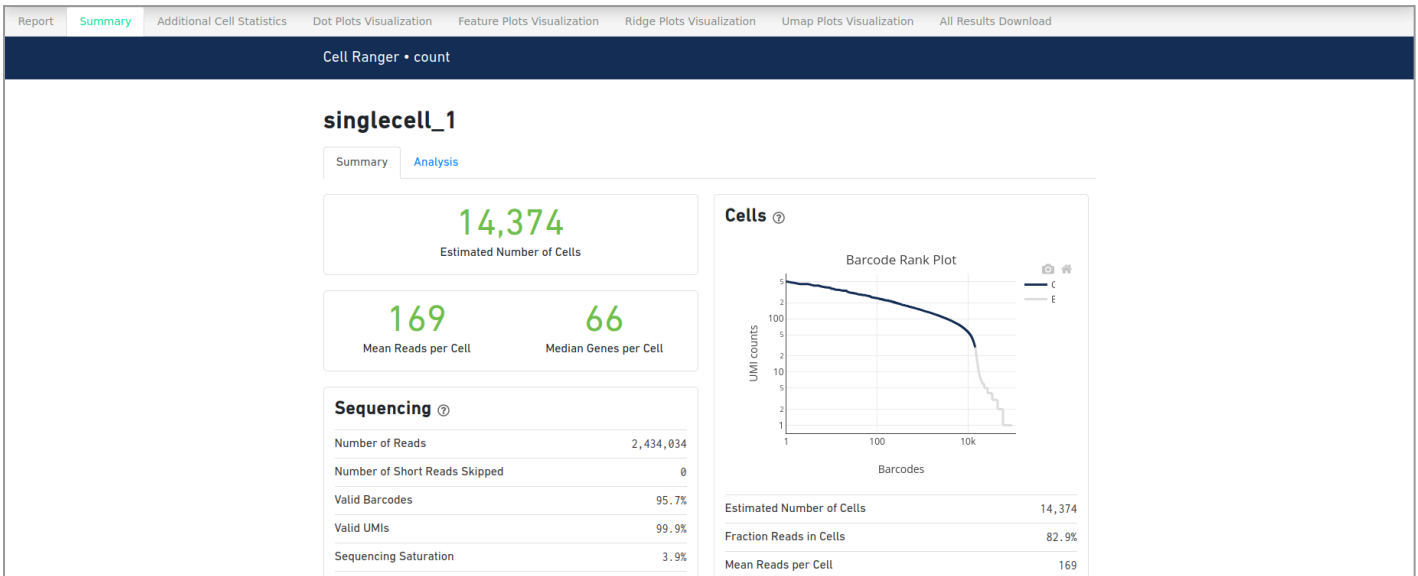
Dotplots Folder: Folder containing all the dotplots needed. (Dotplot_unique, Dotplot_only_specific_genes)

DotPlot_unique.pdf: Top 5 unique differentially expressed genes for each cell cluster based on the p-value and log2fc value.

Dotplot_only_specific_genes.pdf: Same dotplot as the previous ones but for the specific genes you selected.

You can find all these files compressed at their respective zip file.

In the Report tab you can see a generated report that includes descriptions for every step and every parameter of the single-cell RNA-Seq Differential Expression Pipeline for your job.



In the Summary tab you can see a summary of a variety of metrics from the first step of the single-cell RNA-Seq Differential Expression Analysis and some T-SNE plots and information about the Top features by Cluster.

Report	Summary	Additional Cell Statistics	Dot Plots Visualization	Feature Plots Visualization	Ridge Plots Visualization	Umap Plots Visualization	All Results Download				
Total Markers		Markers with Cluster Annotation	Average Expression of genes	Barcode Cluster							
Total Markers Results											
Download Total Markers CSV											
Gene	T1	P value	Average log2 Fold Change	T1	Percentage of cells 1	T1	Percentage of cells 2	T1	Adjusted P value	Cluster	T1
RPL3		5.635e-12	0.704		0.445		0.304		2.063e-07	0	
MT-ATP6		4.861e-10	-0.539		0.451		0.627		1.779e-05	0	
HIST1H4C		4.157e-09	-1.03		0.094		0.21600000000000003		0	0	
TUBA1B		7.684e-09	-0.978		0.094		0.214		0	0	
HSP90AA1		4.382e-08	-0.904		0.10800000000000001		0.228		0.002	0	
MT-CO3		1.562e-07	-0.325		0.6759999999999999		0.8059999999999999		0.006	0	
RPL13		1.858e-07	0.468		0.584		0.516		0.007	0	
S100A4		6.113e-07	-0.636		0.168		0.292		0.022	0	
CFL1		1.608e-06	-0.792		0.081		0.175		0.059	0	
H2AFZ		1.911e-06	-0.804		0.106		0.207		0.07	0	
ACTG1		3.732e-06	-0.51		0.256		0.389		0.137	0	

In the Additional Cell Statistics tab the user can view four different tabs that represent different information for the genes of the input files. The results for these four different tabs can be downloaded at the respective tab. At the Total Markers tab, markers (differentially expressed genes) and associated statistics (p-values, average log2 Fold change etc) can be found.

Total Markers	Markers with Cluster Annotation	Average Expression of genes	Barcode Cluster						
Markers with Cluster Annotation Results									
Download Markers with Cluster Annotation CSV									
Cluster	T1	Cell type	T1	Cell type score	T1	Cell type related markers	T1	PMID	T1
RPL3, MT-ATP6, HIST1H4C, TUBA1B, HSP90AA1, MT-CO3, RPL13, S100A4, CFL1, H2AFZ, ACTG1, TMSB4X, EEF1A1, RPL41, RPS15A, FTL, RPL32, RPS17, LGALS1, HMGB2, RPL39, RPS15, RPS4X, HINT1, UBE2S, RPL34, IFI27, RPL36A, SUB1, PPN1, RPL18, MT-ND5, RPS6, HNRNP2B1, COTL1, S100A6, TRAC, HMGB1, TXN, RPL29, S100A11, RPS2, TPI1, RPL14, SNHG29, RPL28, TUBB, BNIP3, ACTB, VIM, MYL12A		Dendritic Cell		0.65		FTL, S100A11, S100A4, TXN		28428369.0	
UBE2C, CALM2, UBE2S, TUBA1B, TUBB, ARL6IP1, PTGES3, CKS2, H2AFZ, ACTG1, GNG5, HNRNP3, LGALS1, HMGB1, STMN1, HMGB2, EEF1A1, TUBA4A, CALM3, JPT1, HIST1H4C, HNRNP2B1, TXNIP, RPS15, RPS18, RPL21, PSME1, STAT1, NUCKS1, RPS9, EEF1G, RPL12, COX8A, UBB, RPL13, ATP5IF1, RPS27L, MYL12B, TM6SF2, RPL3, H3F3B, RBX1, FTH1, MT2A, RPL18, RPL8, S100A4		Dendritic Cell		0.65		FTH1, MT2A, S100A4, TXNIP, STMN1		28428369.0	
SERBP1, S100A4, PRELID1, NACA, NPM1, ATP5F1C, ZFAS1, SEC61G, COX7A2, RPL12, DBI, NDUFA4, EEF1A1, PPIB, NUCKS1, NCL, BDNF, DUT, UOCHR, RPS3A, SLC25A6, UBALD2, COX8A, RPL18A, CLIC1, RPL6, GSTP1, PSME2, ATP5MG, TRAC, COX6B1, PARK7		Plasmacytoid Dendritic Cell		0.61		PARK7, SEC61G		28428369.0	

At the Markers with Cluster Annotation tab, the results of the Cluster Annotation step can be found.

Total Markers Markers with Cluster Annotation **Average Expression of genes** Barcode Cluster

Average Expression of genes Results

Below you can see the first 500 rows of the generated Average Expression of genes csv. You can download the full results by clicking the "Download Average Expression of genes CSV" button.

[Download Average Expression of genes CSV](#)

Gene	Dendritic Cell_0	Dendritic Cell_1	Plasmacytoid Dendritic Cell_2	Dendritic Cell_3	Dendritic Cell_4	NA_5	Activated T Cell_6	Dendritic Cell_7	Dendritic Cell_8
MT-C01	135.076	145.658	145.848	150.907	150.987	156.787	145.734	155.693	163.858
MALAT1	152.296	161.678	136.656	123.996	133.849	138.266	146.356	139.589	155.297
TMSB4X	118.167	151.634	131.086	128.572	143.985	155.222	139.74	133.718	141.753
MT-C02	125.808	123.558	128.09	143.324	137.983	127.243	126.002	137.135	136.51
B2M	116.047	128.588	135.37	125.947	129.42	139.375	139.114	129.495	128.61
MT-C03	92.926	118.959	116.513	112.321	100.728	129.181	118.24	132.678	116.888
TMSB10	98.179	97.026	97.815	75.135	106.436	96.976	112.073	107.371	85.869
MT-ATP6	47.766	60.765	67.006	80.226	69.105	77.966	68.456	61.039	83.244
MT-ND4	69.709	59.943	63.879	61.415	73.183	63.785	69.906	62.776	74.048
RPS18	74.245	53.424	65.431	64.048	64.979	64.94	70.533	71.907	66.842
RPL41	72.908	52.448	56.155	50.38	61.734	51.419	63.918	62.974	63.469
RPL28	66.65	51.777	57.414	56.278	54.845	63.33	50.525	50.247	67.303
RPLP1	57.212	53.647	58.495	56.696	59.822	54.12	55.352	58.314	62.979

At the Average Expression of genes Results tab the first 500 rows of the generated Average Expression of genes file can be found and it contains the expression levels of every gene for every cluster.

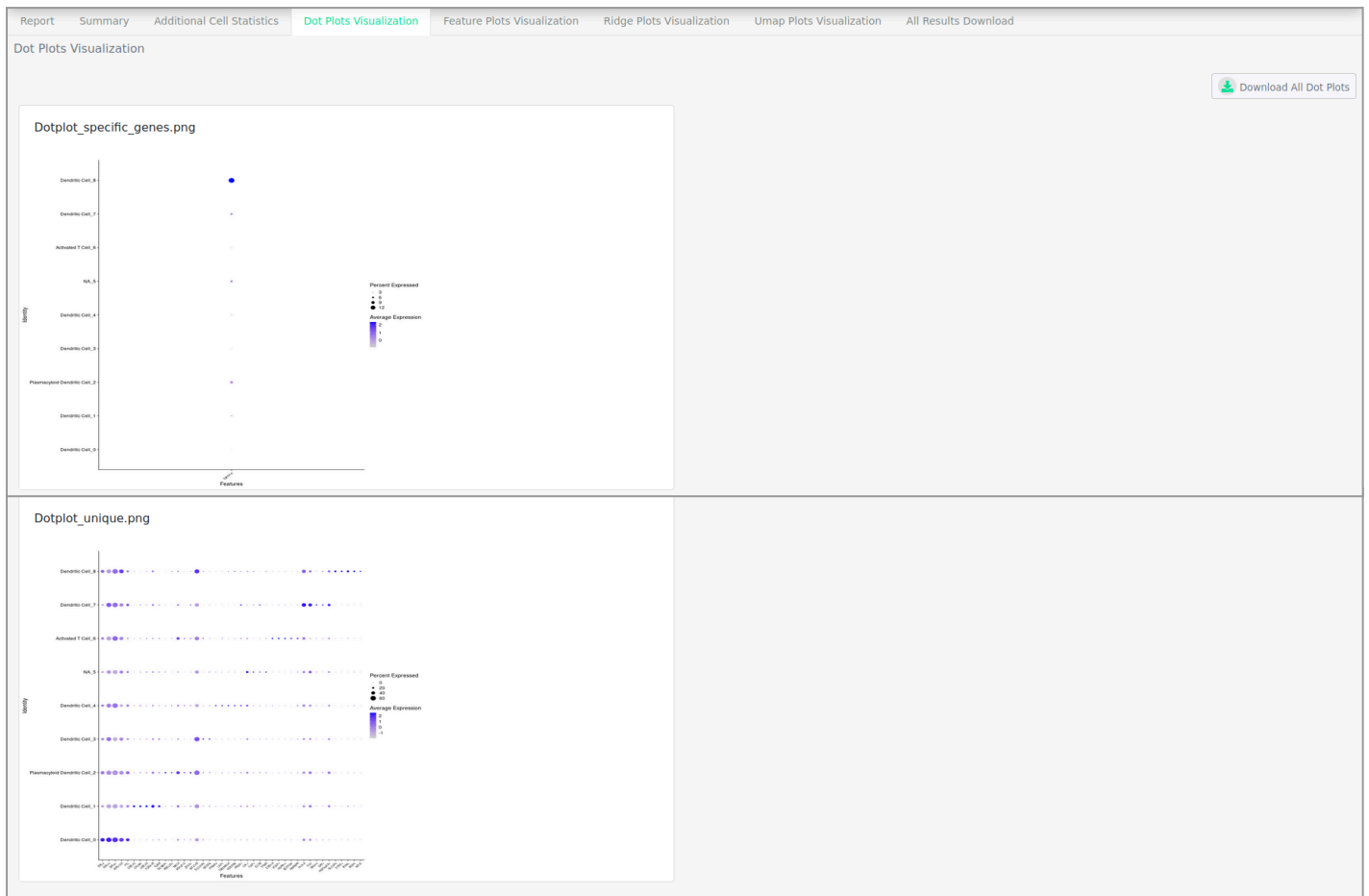
Total Markers Markers with Cluster Annotation Average Expression of genes **Barcode Cluster**

Barcode Cluster Results

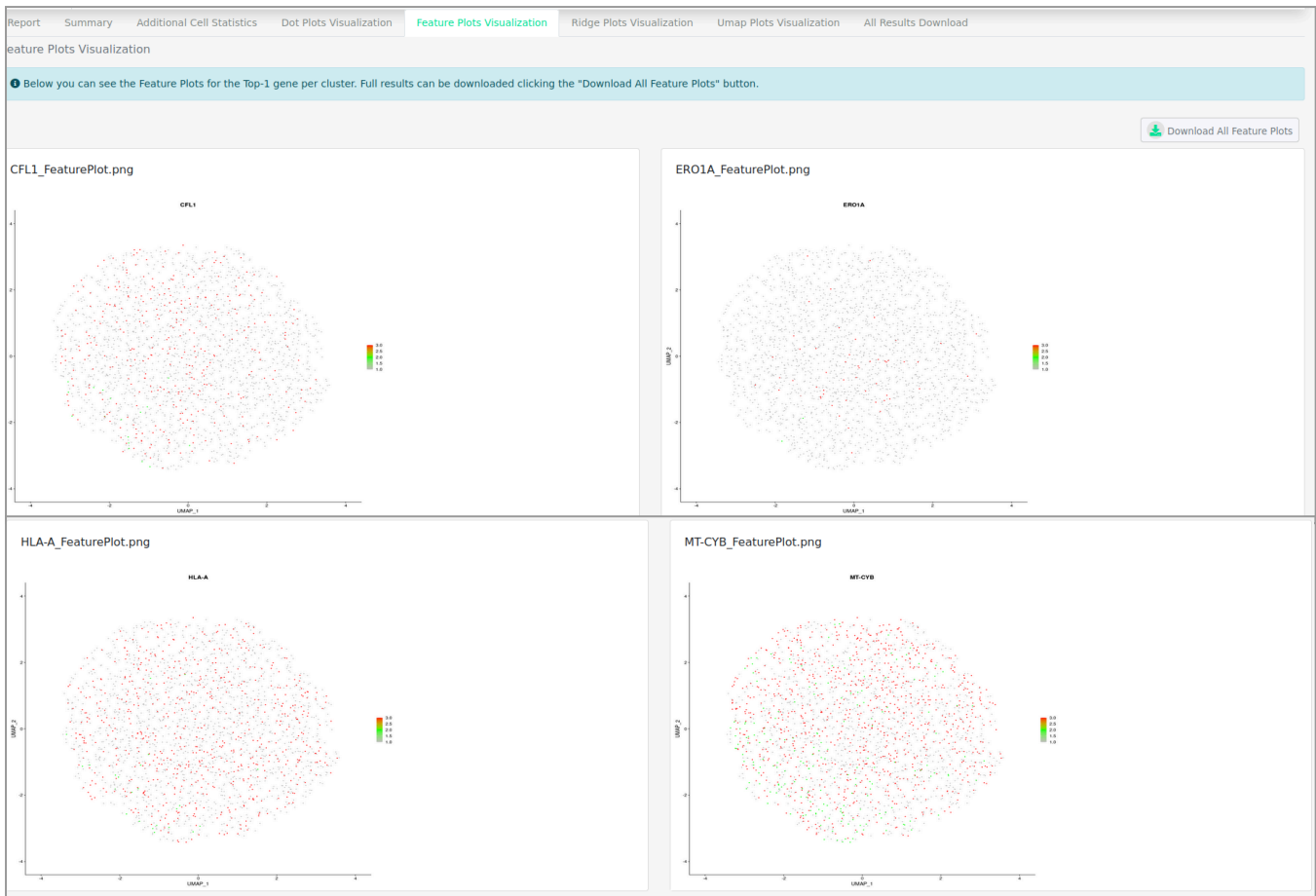
[Download Barcode Cluster CSV](#)

Barcode	Cluster
"AAACCCACATATAGCC-1"	"Activated T Cell_6"
"AAACCCATCAGTCCT-1"	"Dendritic Cell_0"
"AAACCCATCGATGAT-1"	"Dendritic Cell_3"
"AAACGAACAATAGGAT-1"	"Plasmacytoid Dendritic Cell_2"
"AAACGAACACAAGTA-1"	"NA_5"
"AAACGAACATCTATCT-1"	"Dendritic Cell_3"
"AAACGCTAGCTACTGT-1"	"Dendritic Cell_0"
"AAACGCTCAGATCCAT-1"	"Plasmacytoid Dendritic Cell_2"
"AAACGCTTCCATCAGA-1"	"Dendritic Cell_1"
"AAAGAACCATGGCCAC-1"	"Dendritic Cell_1"
"AAAGAACTGCCGATG-1"	"Dendritic Cell_8"
"AAAGGATAGTACAGCG-1"	"Activated T Cell_6"
"AAAGGATCACGAGAAC-1"	"Dendritic Cell_4"
"AAAGGATCACTCATAG-1"	"Dendritic Cell_8"
"AAAGGATGTGCTATA-1"	"Dendritic Cell_1"

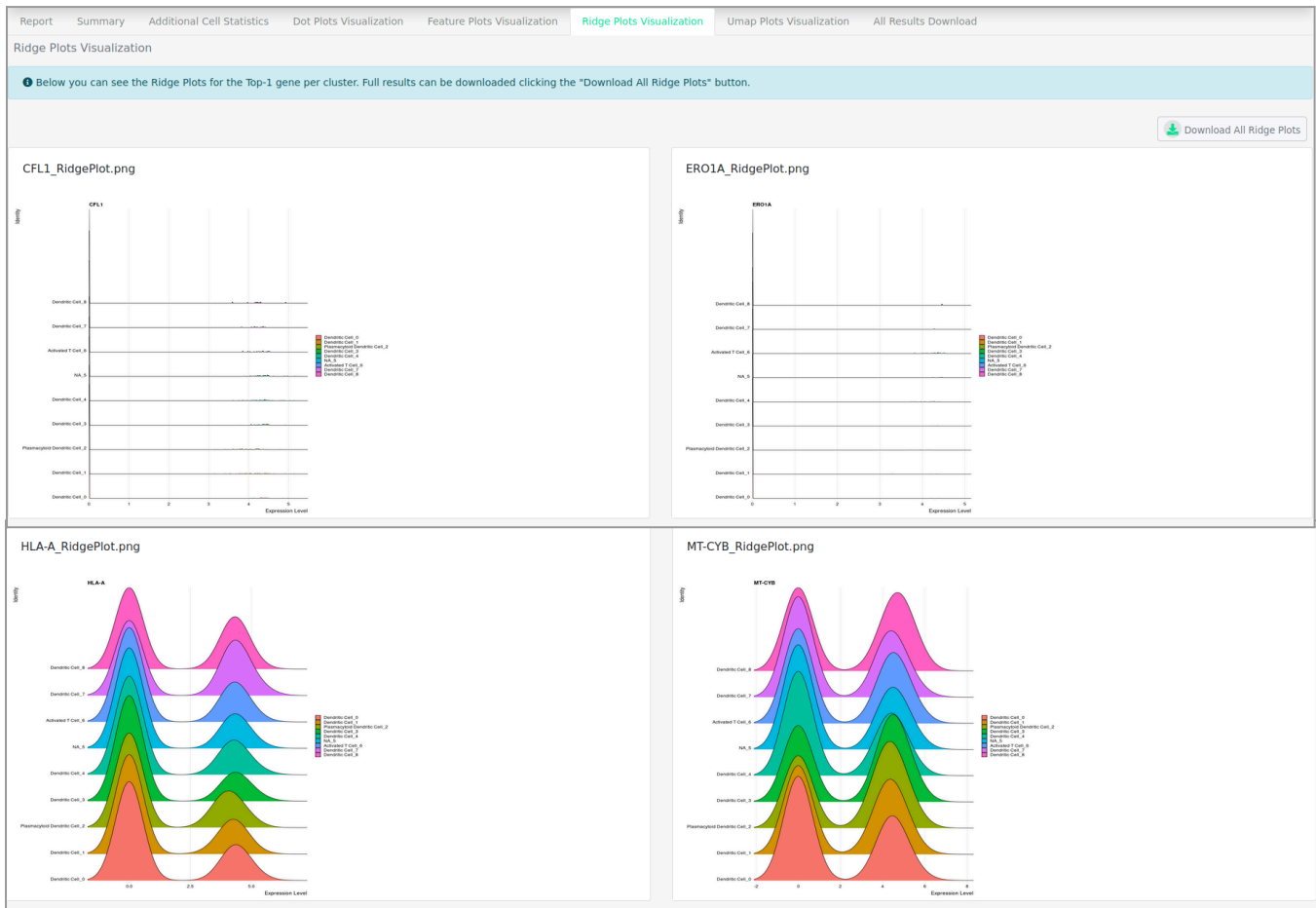
At the Barcode Cluster tab, the Barcode-Cluster matrix can be found.



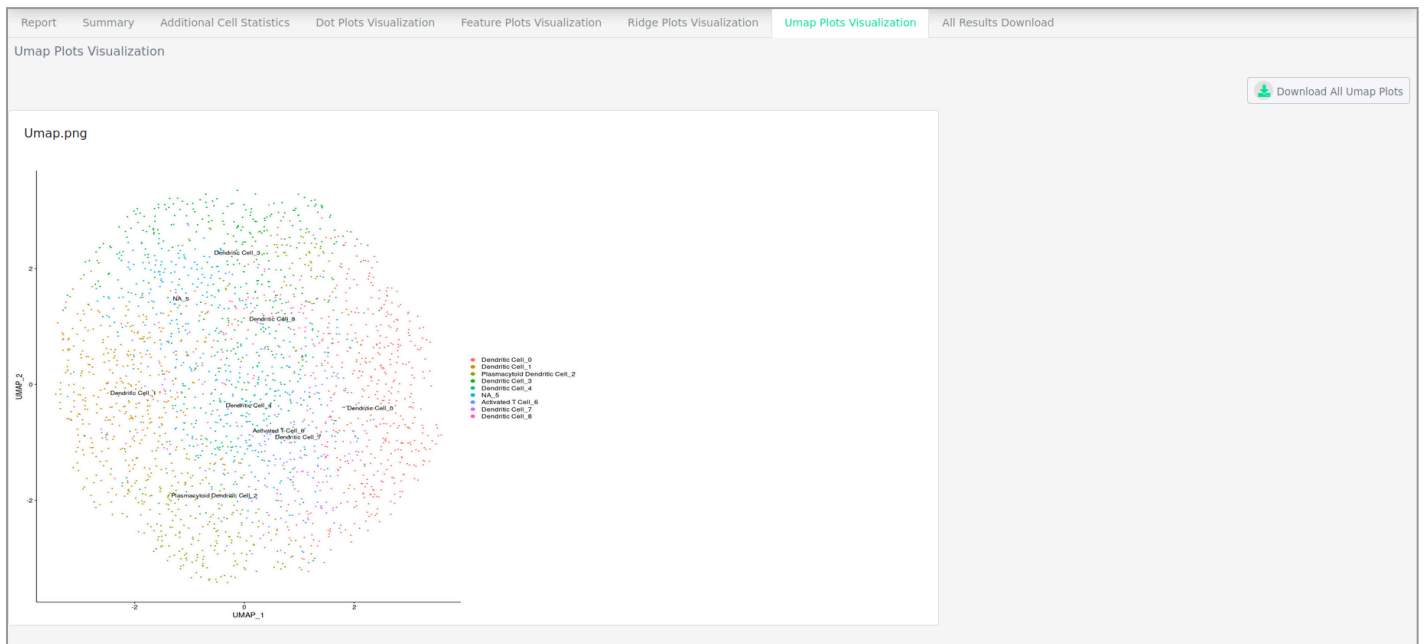
At the Dot Plots Visualization tab you can see the two Dot plots that are created. The first one is a Dot Plot with only the genes you specified at the manual parameters and the second one is a Dot Plot that shows the Top 5 unique differentially expressed genes for each cell cluster based on the p-value and log₂ fold change value. These plots can be downloaded.



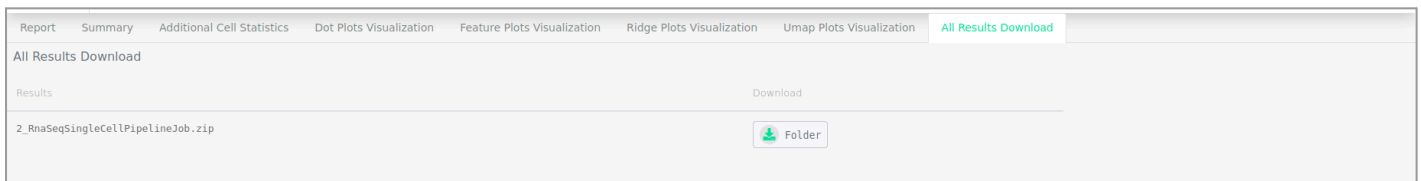
At the Feature Plots Visualization tab the Feature Plots for the Top-1 gene per cluster can be found. The Feature Plots of all the genes can be downloaded.



At the Ridge Plots Visualization tab the Ridge Plots for the Top-1 gene per cluster can be found. The Ridge Plots of all the genes can be downloaded.



At the Umap Plots Visualization tab the Umap Plots can be found. The Umap Plot can be downloaded.



At the All Results Download tab, all the results of your job can be downloaded.

Deconvolve Data against single-cell RNA-seq Analysis

You can deconvolve data against a single-cell RNA-Seq dataset. Firstly, it is required to import the single-cell RNA-Seq 10x datasets, the Matrix, the Feature and the Barcodes datasets. Secondly, you must import the biomarker files, a BulkRnaSeq file, the Biomarkers Labels and the Barcode-Cluster file. This Pipeline uses the SCDC method (Bulk Gene Expression Deconvolution by Multiple Single-Cell RNA Sequencing Referencing) to perform the deconvolution.

To start the deconvolution pipeline:

Click in the menu “InSyBio ncRNASeq” → “single-cell RNA-Seq Data Analysis” → “single-cell RNA-Seq Pipeline Dashboard”, select the “Add new job” button and then choose the “Deconvolve Data against single-cell RNA-seq Analysis” option. Then do the following steps:

- Upload your files of single-cell RNA-Seq Experiments Matrix, Features and Barcodes datasets.
- Upload your fastq or Read Count Biomarker files.

InSyBio Suite - Deconvolve Data Against Single-cell RNA-Seq Analysis

☰
🔒
🔔

InSyBio Beta User

🗨
?

[← Dashboard](#)

scRNAseq Files

* Required information

<p>Matrix Title: *</p> <input style="width: 95%;" type="text"/>	<p>Features Title: *</p> <input style="width: 95%;" type="text"/>	<p>Barcodes Title: *</p> <input style="width: 95%;" type="text"/>
<p>Matrix Filename: *</p> <input style="width: 95%;" type="text"/>	<p>Features Filename: *</p> <input style="width: 95%;" type="text"/>	<p>Barcodes Filename: *</p> <input style="width: 95%;" type="text"/>
<div style="background-color: #00b050; color: white; padding: 2px 5px; border-radius: 3px; display: inline-block;"> Select from Data Store </div>	<div style="background-color: #00b050; color: white; padding: 2px 5px; border-radius: 3px; display: inline-block;"> Select from Data Store </div>	<div style="background-color: #00b050; color: white; padding: 2px 5px; border-radius: 3px; display: inline-block;"> Select from Data Store </div>
<div style="background-color: #0070c0; color: white; padding: 2px 5px; border-radius: 3px; display: inline-block;"> Upload to Data Store </div>	<div style="background-color: #0070c0; color: white; padding: 2px 5px; border-radius: 3px; display: inline-block;"> Upload to Data Store </div>	<div style="background-color: #0070c0; color: white; padding: 2px 5px; border-radius: 3px; display: inline-block;"> Upload to Data Store </div>

Biomarker Files

BulkRNAseq File *

Title:

Filename:

Select file from Data Store

Go to Data Store to Upload File

Biomarkers Labels *

Title:

Filename:

Select file from Data Store

Go to Data Store to Upload File

Barcode Cluster *

Title:

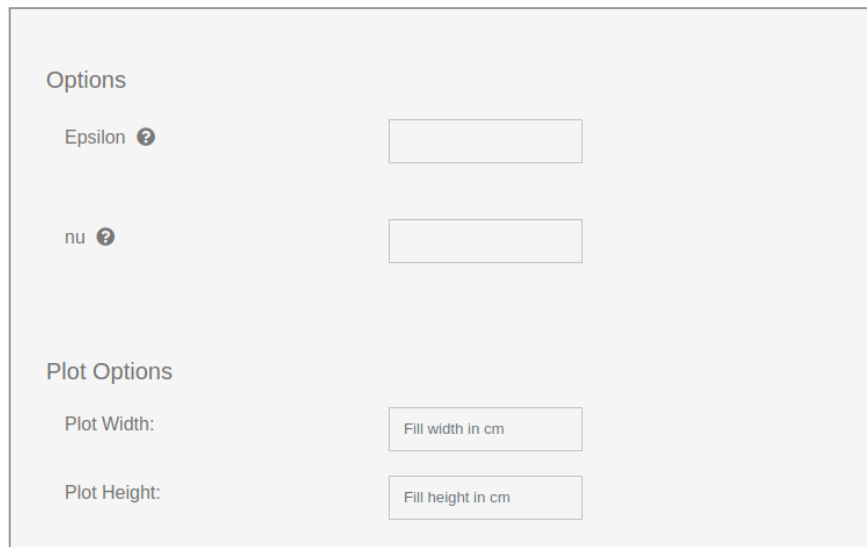
Filename:

Select file from Data Store

Go to Data Store to Upload File

- Fill in the Epsilon integer, a small constant number used for convergence criteria.

- Fill in the nu integer, a small constant number to facilitate the calculation of variance.
- Fill in the Plot options
 - Plot width
 - Plot height



The screenshot shows a form titled 'Options' with two input fields: 'Epsilon' and 'nu', each with a help icon. Below this is a section titled 'Plot Options' with two input fields: 'Plot Width' and 'Plot Height', each with a placeholder text 'Fill width in cm' and 'Fill height in cm' respectively.

- Submit your job pressing the respective button.

To view the results:

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

Status	Job ID	Job Type	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step	Actions
Completed	21	RNASeq Single Cell Velocity Analysis		11/26/73, 3:22 AM	1/16/24, 1:59 PM	-	Secondary Single Cell Analysis	View Results
Completed	20	Deconvolve Data against single-cell RNA-seq Analysis		8/12/11, 6:41 AM	1/15/24, 2:14 PM	1/15/24, 2:15 PM		View Results
Completed	19	RNASeq Single Cell Velocity Analysis		3/29/80, 9:11 PM	1/12/24, 8:51 AM	1/12/24, 9:06 AM	Secondary Single Cell Analysis	View Results
Completed	18	Cell Chat Analysis		12/19/68, 6:13 AM	1/12/24, 8:29 AM	1/12/24, 8:33 AM	Single Cell Alignment	View Results
Completed	17	RDS Conversion		2/9/59, 5:58 AM	1/11/24, 11:16 AM	1/11/24, 11:18 AM	Secondary Single Cell Analysis	View Results

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 that they were produced.

In the Results CSV tab, you can see the three generated csv files, basic.csv, yhat.csv and props1.csv. Props1 shows the predicted proportions of cell clusters in every sample, Yhat shows the predicted proportions of cell clusters in every gene and Basis represents the basis matrix.

At the Plots Visualization tab you can see the plot that is created. This plot represents the predicted proportions of cell clusters in every sample. This plot can be downloaded.

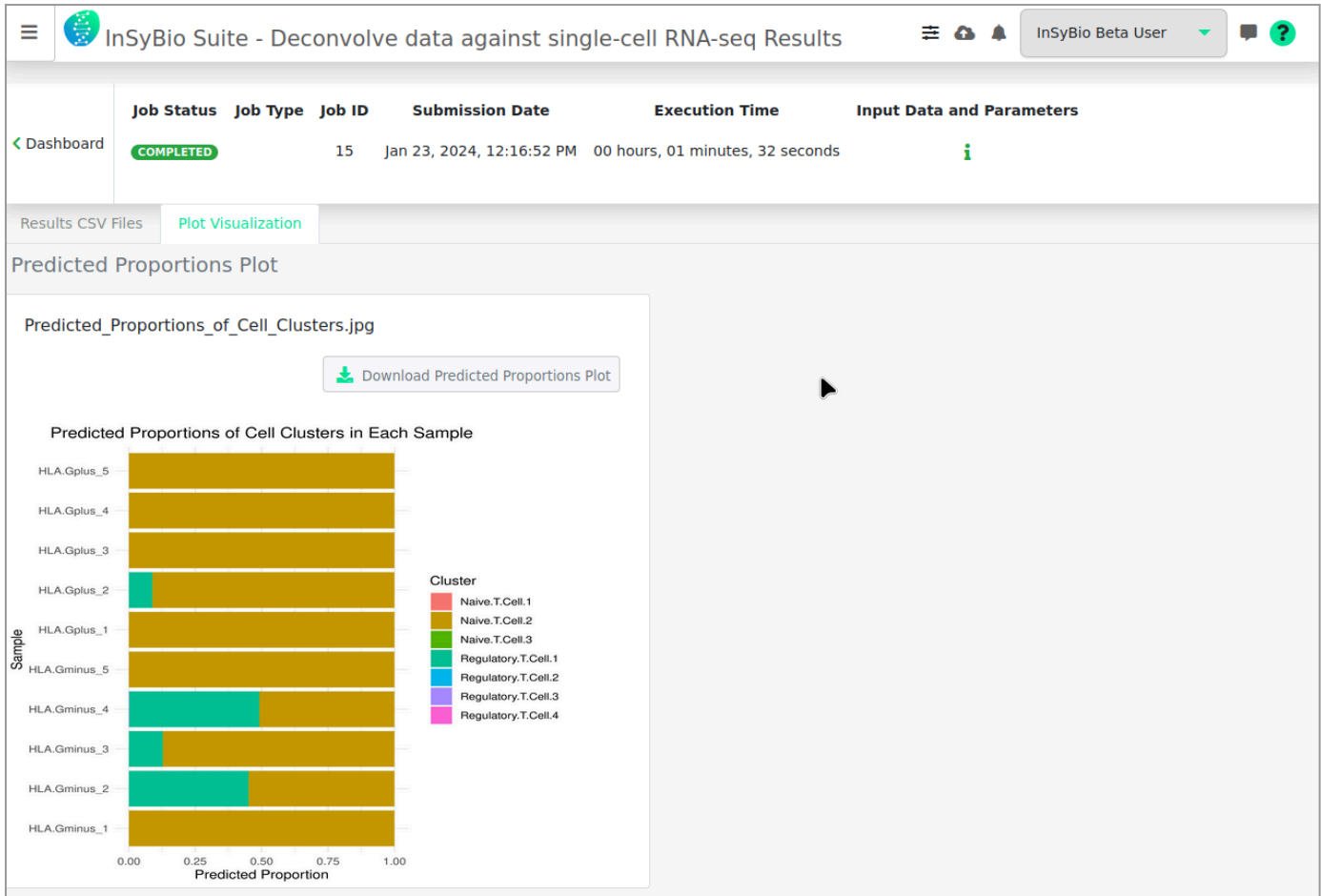
InSyBio Suite - Deconvolve data against single-cell RNA-seq Results

InSyBio Beta User

Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
COMPLETED		15	Jan 23, 2024, 12:16:52 PM	00 hours, 01 minutes, 32 seconds	i

Results CSV Files | Plot Visualization

CSV File	Download
Basis CSV	Download Basis CSV
Props CSV	Download Props CSV
Yhat CSV	Download Yhat CSV



Velocity single-cell Analysis

You can do the Velocity single-cell Analysis. Firstly, it is required to import the single-cell RNA-Seq 10X datasets, the Matrix, the Feature and the Barcodes datasets. This Pipeline uses the velocity tool to estimate the RNA velocities of single-cells and the monocle3 and scvelo packages to identify trajectories and further analyse the estimated velocities.

To start the Velocity single-cell Analysis:

Click in the menu “InSyBio ncRNASeq” → “single-cell RNA-Seq Data Analysis” → “single-cell RNA-Seq Pipeline Dashboard”, select the “Add new job” button and then choose the “Velocity single-cell Analysis” option. Then do the following steps:

- Upload your files of single-cell RNA-Seq Experiments 10X Matrix, Features and Barcodes datasets.
- Select the transcriptome the input files belong to from our 3 built-in options (Human, Mouse, Human-mouse mixture).
- Select the computation type of velocity.
- Fill in the root nodes, because you need to specify the start of the trajectory, meaning the group (cluster) of cells which is undifferentiated at the beginning of the analysis.

The screenshot displays the 'InSyBio Suite - Velocity single-cell Analysis' web interface. At the top, the user is logged in as 'InSyBio Beta User'. The main form is divided into three columns for Matrix, Features, and Barcodes data. Each column contains a title field and a filename field, both with 'Select from Data Store' and 'Upload to Data Store' buttons. Below these is a 'Bam File' section with 'Title' and 'Filename' input fields, and 'Select file from Data Store' and 'Go to Data Store to Upload File' buttons. A '* Required information' note is visible in the top right corner of the form area.

Transcriptome: Human

Computation of velocity: Stochastic

Root nodes ? : ex.1,2,3 or if they have annotated clusters T Regulatory Cell 1, T Regulatory Cell 2, Naive Cell 3.

Cluster annotation

Species: --Select Action--

Tissue ? : --Select Action--

- Select if you want to manually configure other parameters of the job. If you don't, our Default Options will be applied. Possible manual options are:
 - First filtering:
 - Minimum cells
 - Minimum features
 - Secondary filtering:
 - nFeature_RNA with lower and upper limits
 - nCount_RNA with lower and upper limits
 - Feature Extraction Method
 - Shared Nearest Neighbor (SNN) Graph
 - K parameter (k-nearest- neighbor)
 - Clustering
 - Resolution parameter
 - Threshold (logfc)
 - Minimum Pct

Advanced Options +

First filtering

Minimum cells:

Minimum features:

Secondary filtering

nFeature_RNA ? :

Lower limit:

Upper limit:

nCount_RNA ? :

Feature Extraction Method

Shared Nearest Neighbor (SNN) Graph

k parameter (k-nearest-neighbor):

Clustering

Resolution parameter ? :

- Submit your job pressing the respective button.

To view the results:

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

The screenshot shows the 'InSyBio Suite - Single Cell RNA-Seq Differential Expression Pipeline Dashboard'. On the left is a navigation menu with options like 'InSyBio Interact', 'InSyBio ncRNASeq', 'InSyBio Bionets', 'InSyBio Biomarkers', 'InSyBio DNA-Seq', 'InSyBio Pipelines', and 'InSyBio DataStore'. The main area features a '+ Add new Job' button and a 'Filter Jobs' dropdown set to 'Show All'. A summary bar indicates 13 Completed, 1 Running, 0 Pending, and 4 Error jobs. Below is a table with columns for Status, Job ID, Job Type, Input File(s), Submission Date, Start Execution Date, Completion Date, Current Step, and Actions.

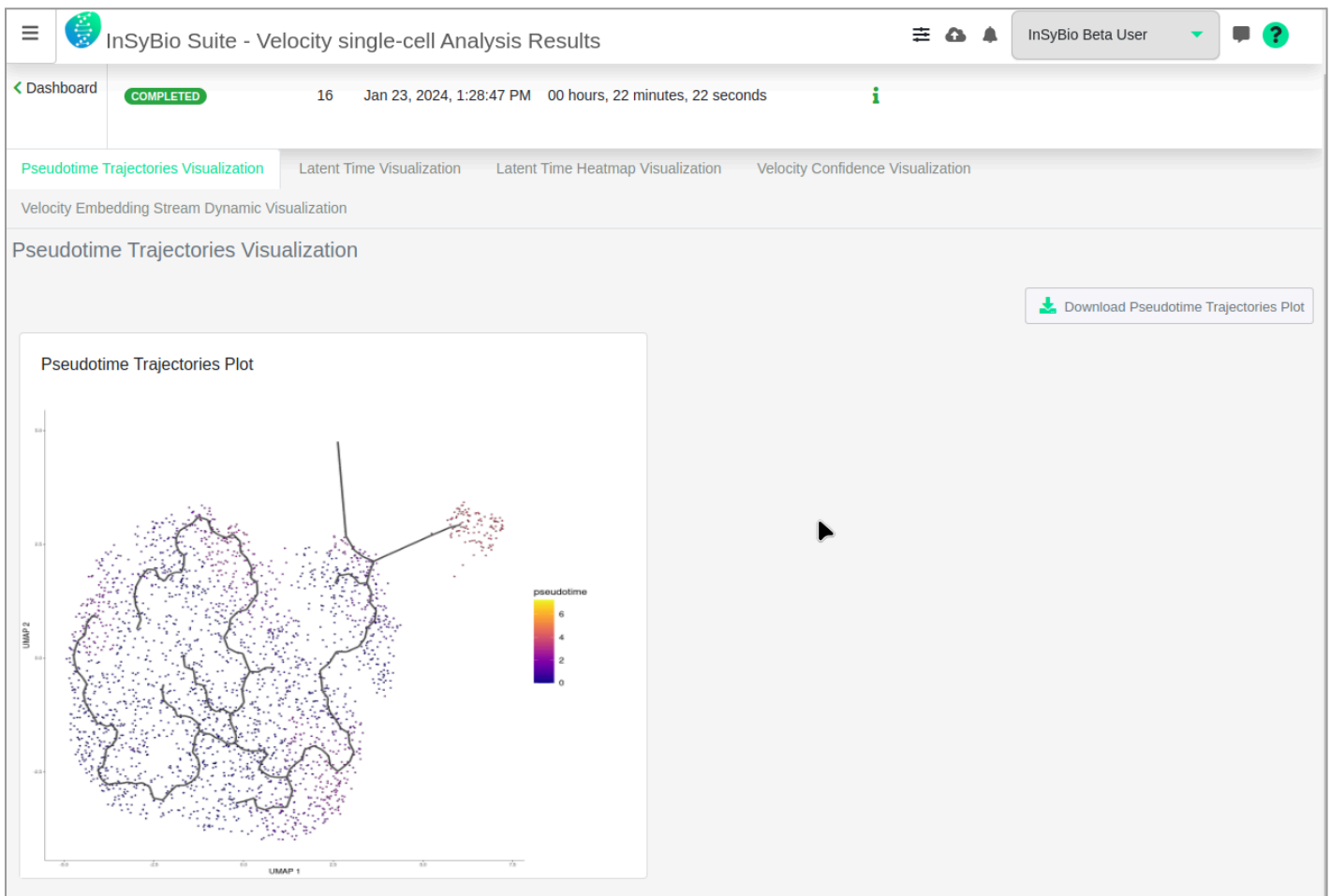
Status	Job ID	Job Type	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step	Actions
Completed	21	RNASeq Single Cell Velocity Analysis		11/26/73, 3:22 AM	1/16/24, 1:59 PM	-	Secondary Single Cell Analysis	View Results
Completed	20	Deconvolve Data against single-cell RNA-seq Analysis		8/12/11, 6:41 AM	1/15/24, 2:14 PM	1/15/24, 2:15 PM		View Results
Completed	19	RNASeq Single Cell Velocity Analysis		3/29/80, 9:11 PM	1/12/24, 8:51 AM	1/12/24, 9:06 AM	Secondary Single Cell Analysis	View Results
Completed	18	Cell Chat Analysis		12/19/68, 6:13 AM	1/12/24, 8:29 AM	1/12/24, 8:33 AM	Single Cell Alignment	View Results
Completed	17	RDS Conversion		2/9/59, 5:58 AM	1/11/24, 11:16 AM	1/11/24, 11:18 AM	Secondary Single Cell Analysis	View Results

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 that they were produced.

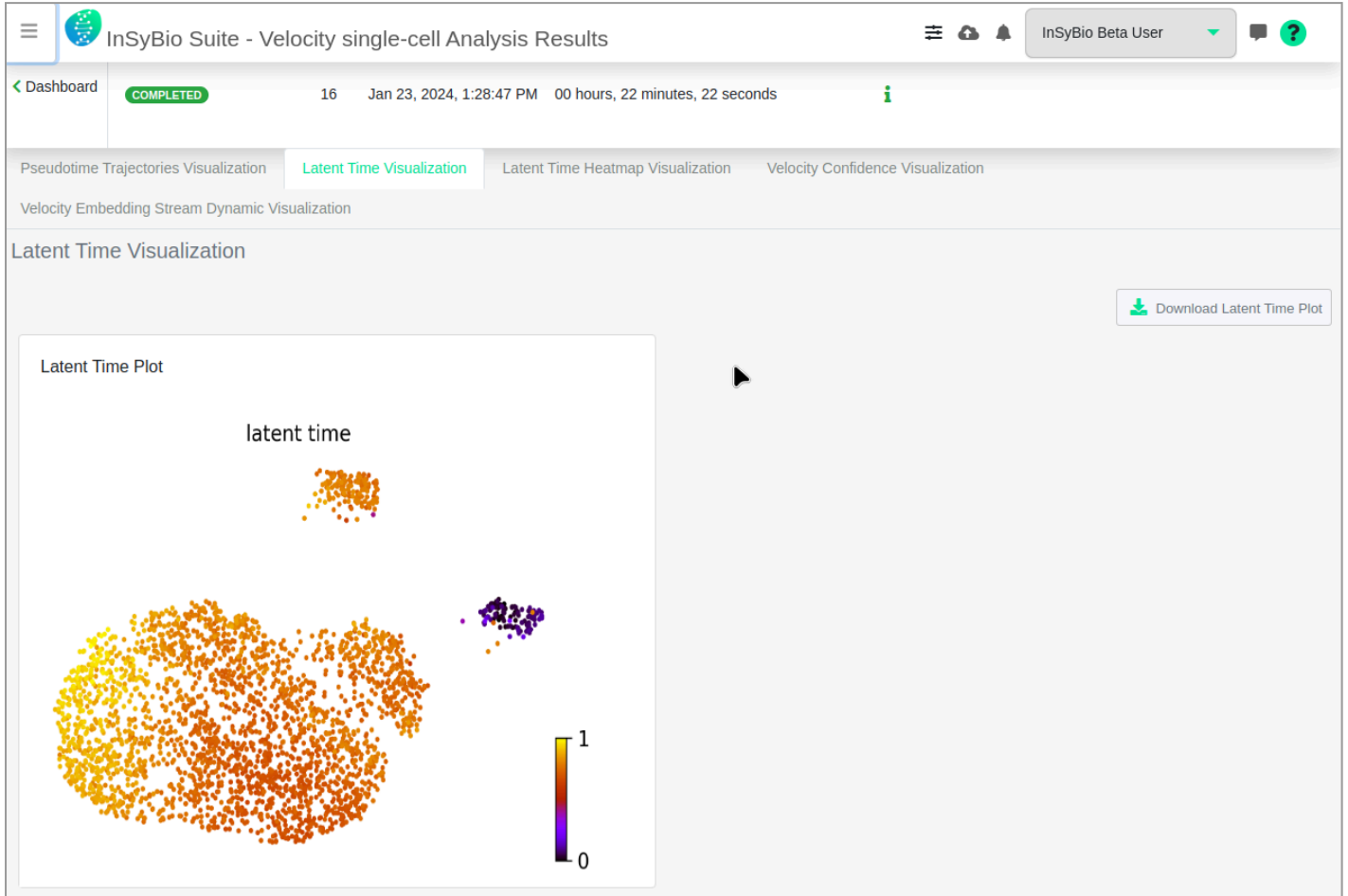
Depending on the computation type of velocity you selected, different tabs will appear.

For dynamic analysis of velocity, five different tabs are present, each one representing a different step in the analysis and a produced plot.

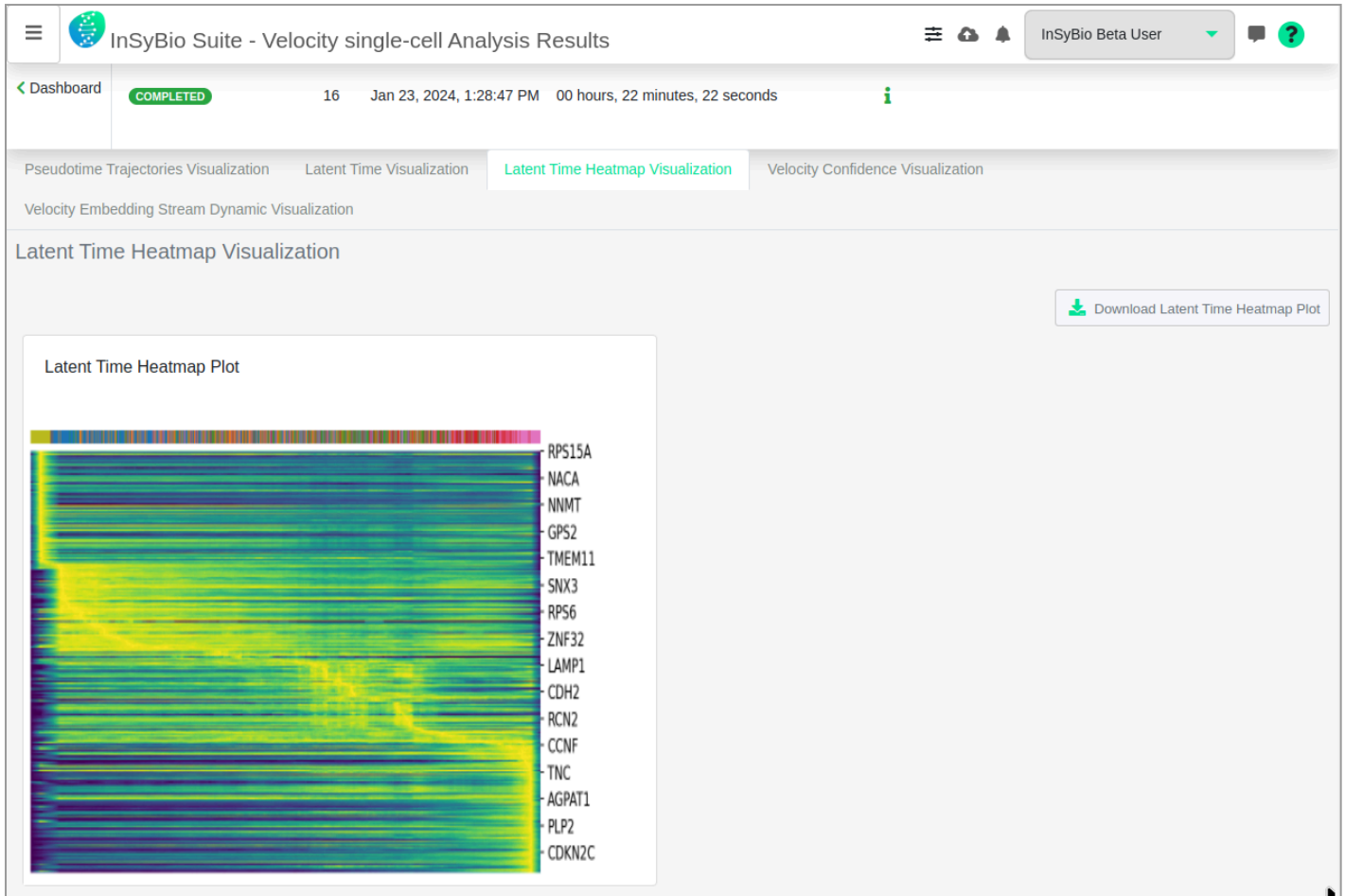
In the Pseudotime Trajectories Visualization tab, the plot visualizes the pseudotime trajectories calculated by monocle3.



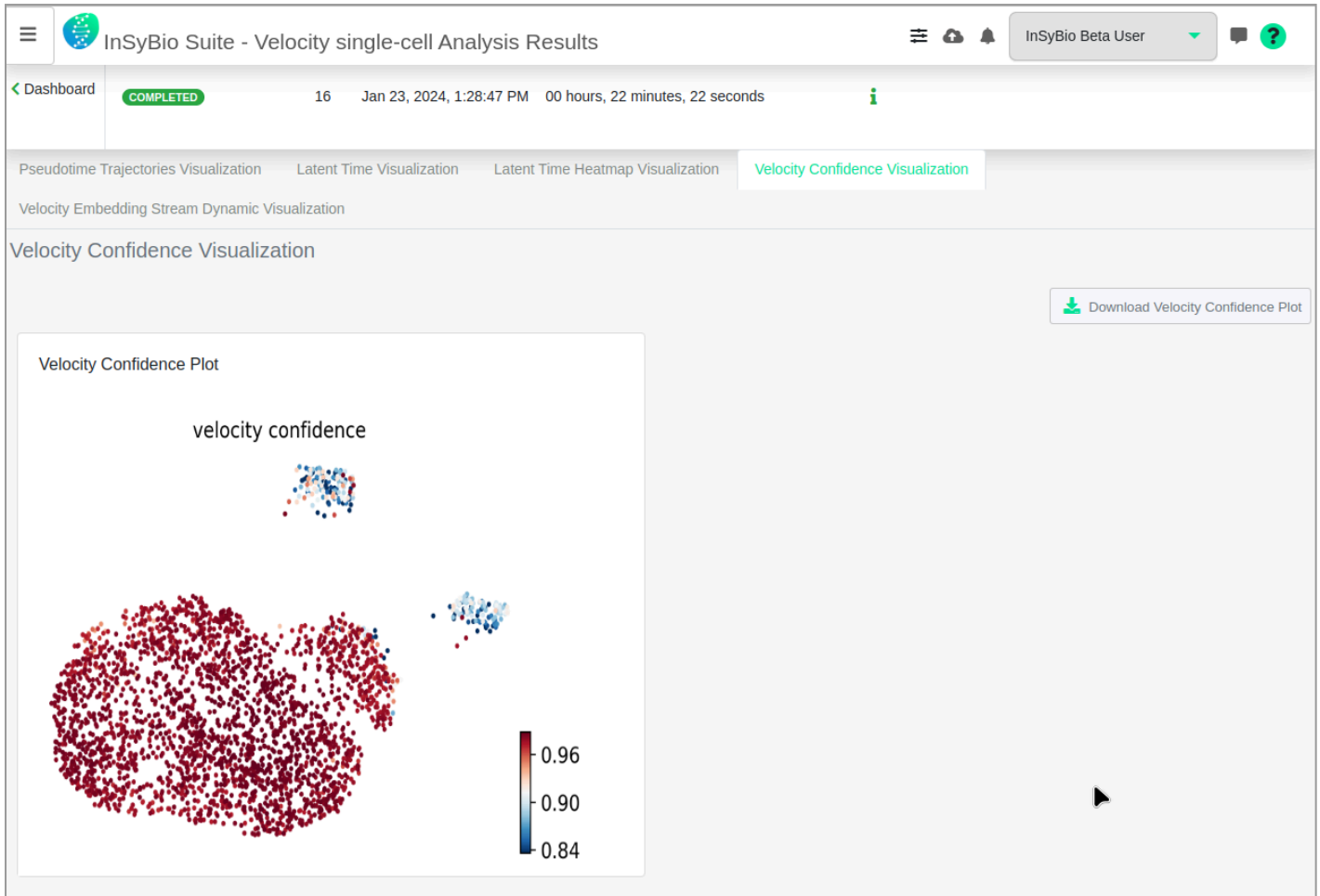
In the Latent time Visualization tab, the plot represents the latent time of the underlying cellular processes, an approximation of the real time experienced by cells as they differentiate.



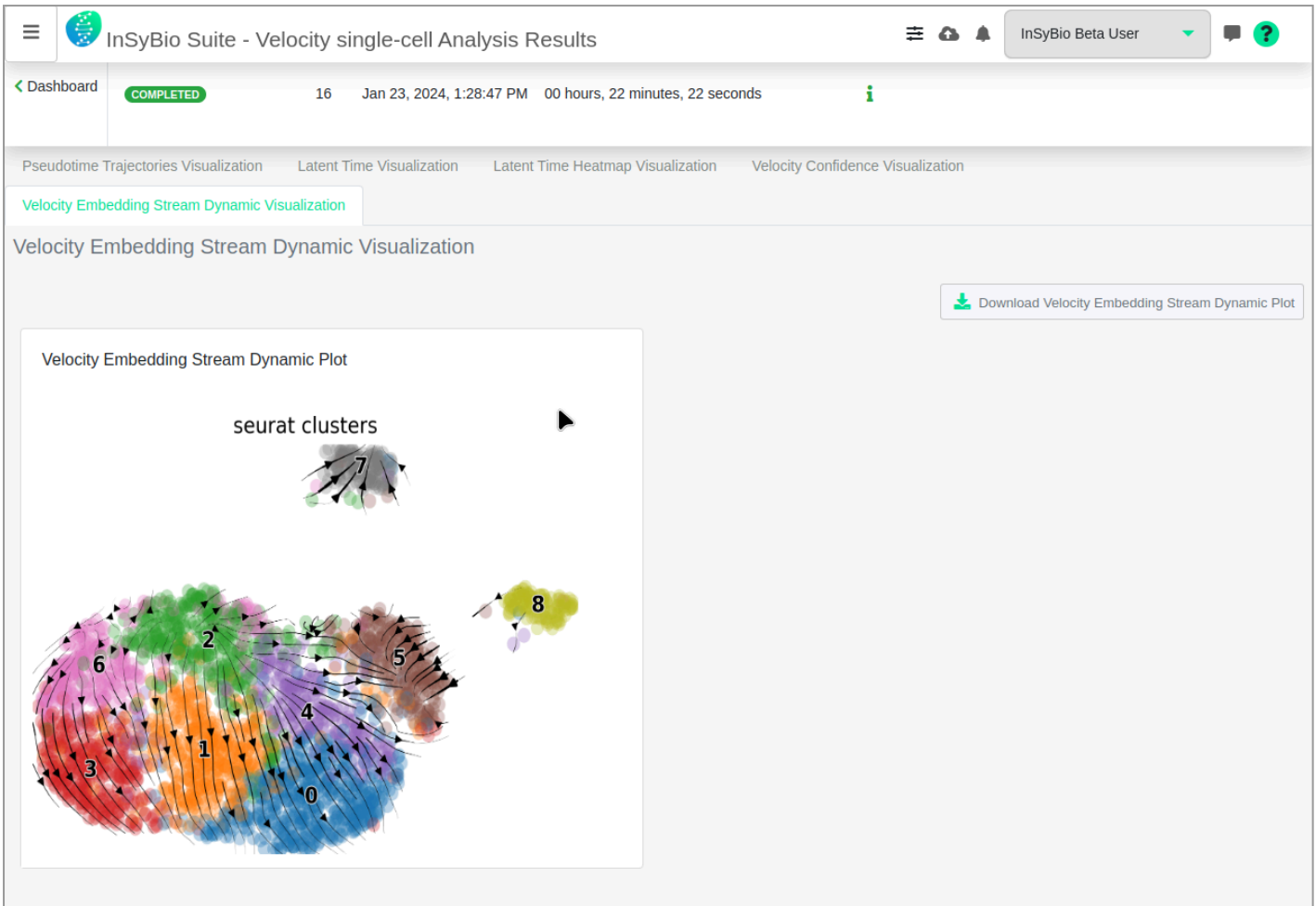
In the Latent time heatmap Visualization tab, the plot represents the latent time heatmap of the top genes.



In the Velocity confidence Visualization tab, the plot represents the computation confidences of velocities.



In the Velocity embedding stream dynamic Visualization tab, the plot visualizes the dynamic stream of velocities.



These plots can also be downloaded individually.

Cell Chat single-cell Analysis

You can do the Cell Chat single-cell Analysis. Firstly, it is required to import the single-cell seurat rds dataset. This pipeline uses the CellChat R toolkit to visualize cell-cell communication from single-cell data.

To start the Cell Chat single-cell pipeline:

Click in the menu “InSyBio ncRNASeq” → “single-cell RNA-Seq Data Analysis” → “single-cell RNA-Seq Pipeline Dashboard” , select the “Add new job” button and then choose the “Cell Chat single-cell Analysis” option. Then do the following steps :

- Upload your seurat object file (.rds format) file, which should already have annotated clusters. These annotations should be accessible by reading the output of the levels function on this object.
- Select if you want to manually configure the plot parameters of the job. If you don't, our Default Options will be applied. Possible manual options are:
 - Plot width
 - Plot height
 - Plot fontsize

InSyBio Suite - Cell Chat single-cell Analysis

Dashboard

RDS File

Title:

Filename:

Select file from Data Store

Go to Data Store to Upload File

Plot Options

Plot Width:

Plot Height:

Font Size:

Submit Job

- Submit your job pressing the respective button.

To view the results:

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

After the analysis, you can select the View Results in the Actions column and view the produced files, that are separated according to the step that they were produced.

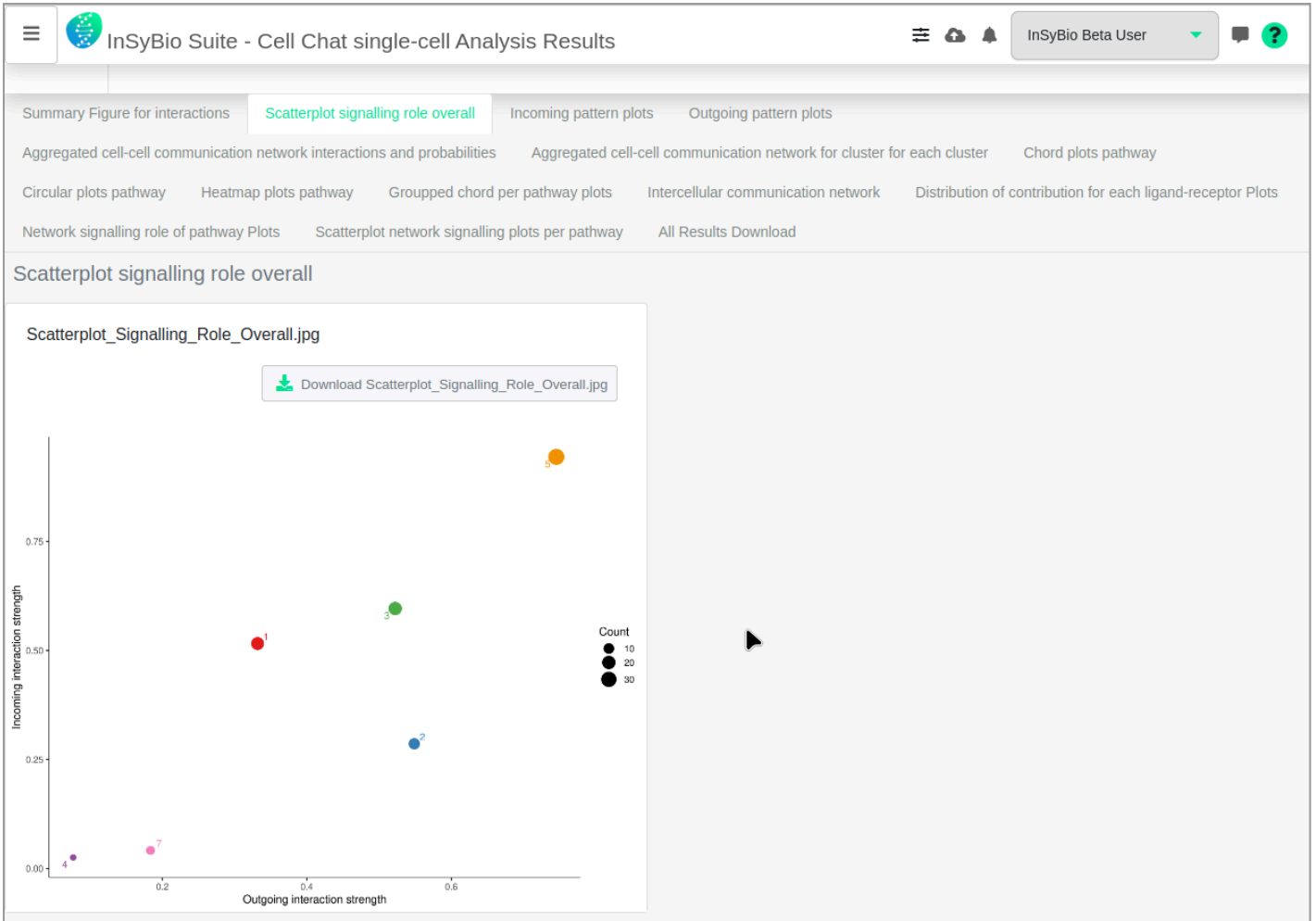
Fourteen different result tabs are present, each of which represents a different analysis performed on the Seurat object. Below a representative example of each tab will be shown.

At the All Results Download tab, all the results of your job can be downloaded.

The screenshot displays the InSyBio Suite interface for 'Cell Chat single-cell Analysis Results'. At the top, the user is identified as 'InSyBio Beta User'. A table shows a job with ID 9, submitted on Jan 23, 2024, at 8:38:31 AM, which has been 'COMPLETED' after an execution time of 00 hours, 11 minutes, and 28 seconds. Below the table, a navigation bar lists various analysis tabs, with 'Summary Figure for interactions' selected. This section contains three pie charts and a 'Download Summary Figure for interactions' button.

Figure ligand_receptor_interaction_database_summary_data.jpg

Chart	Category	Percentage
Chart 1: Secreted Signaling vs. ECM-Receptor vs. Cell-Cell Contact vs. Non-protein Signaling	Secreted Signaling	29.4%
	ECM-Receptor	13.1%
	Cell-Cell Contact	10.8%
	Non-protein Signaling	30.7%
Chart 2: Heterodimers vs. Others	Heterodimers	81%
	Others	9%
Chart 3: KEGG vs. Literature	KEGG	44%
	Literature	9%



☰ InSyBio Suite - Cell Chat single-cell Analysis Results
🔔 InSyBio Beta User

	Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
< Dashboard	COMPLETED		9	Jan 23, 2024, 8:38:31 AM	00 hours, 11 minutes, 28 seconds	i

Summary Figure for interactions Scatterplot signalling role overall **Incoming pattern plots** Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities Aggregated cell-cell communication network for cluster for each cluster Chord plots pathway

Circular plots pathway Heatmap plots pathway Grouped chord per pathway plots Intercellular communication network

Distribution of contribution for each ligand-receptor Plots Network signalling role of pathway Plots Scatterplot network signalling plots per pathway All Results Download

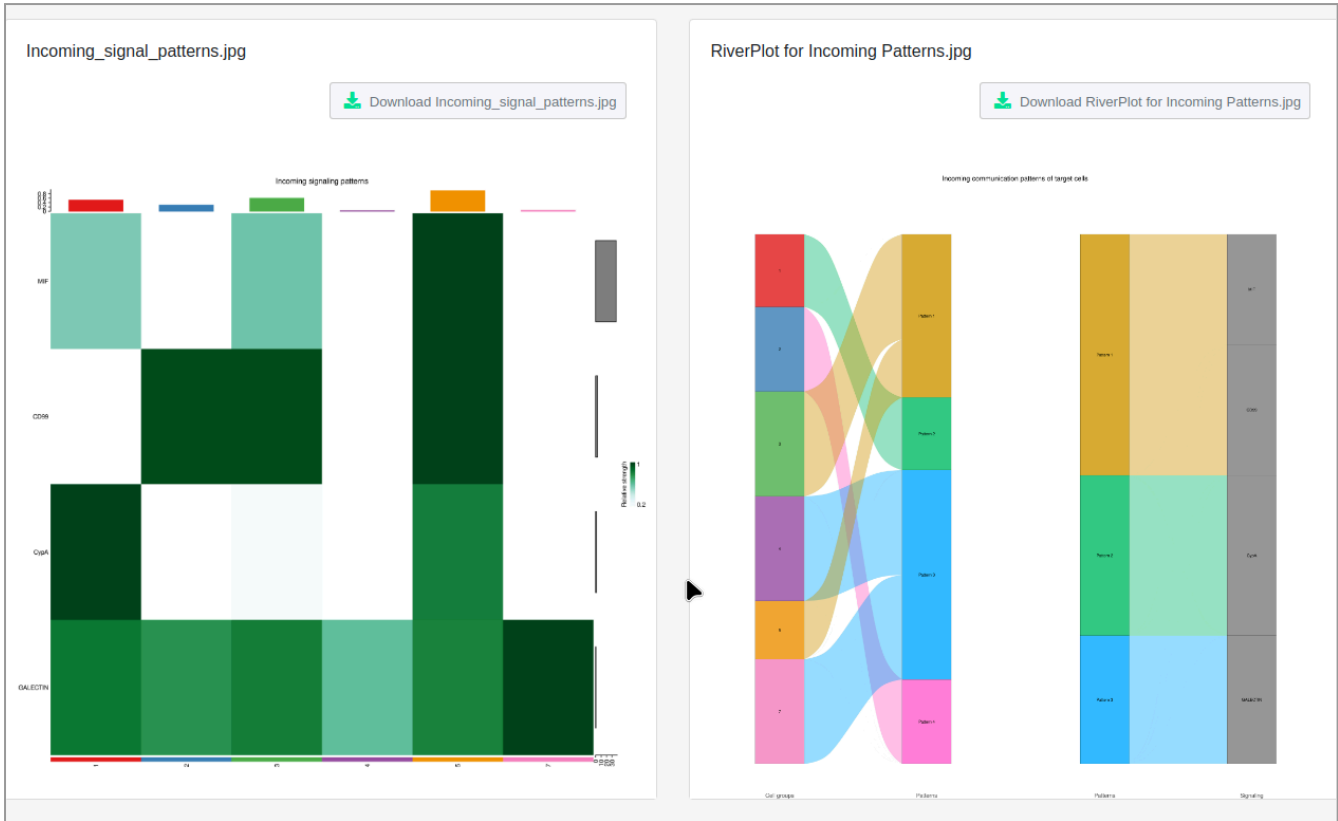
Incoming pattern plots

Dotplot for incoming Patterns.jpg

[Download Dotplot for incoming Patterns.jpg](#)

Incoming Communication Patterns.jpg

[Download Incoming Communication Patterns.jpg](#)



☰ InSyBio Suite - Cell Chat single-cell Analysis Results
☰ 📧 📢 InSyBio Beta User ▼ ?

	Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
← Dashboard	COMPLETED		9	Jan 23, 2024, 8:38:31 AM	00 hours, 11 minutes, 28 seconds	i

Summary Figure for interactions
Scatterplot signalling role overall
Incoming pattern plots
Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities
Aggregated cell-cell communication network for cluster for each cluster
Chord plots pathway

Circular plots pathway
Heatmap plots pathway
Grouped chord per pathway plots
Intercellular communication network

Distribution of contribution for each ligand-receptor Plots
Network signalling role of pathway Plots
Scatterplot network signalling plots per pathway
All Results Download

Aggregated cell-cell communication network interactions and probabilities

Figure Aggregated cell-cell communication network_for cluster_0.jpg

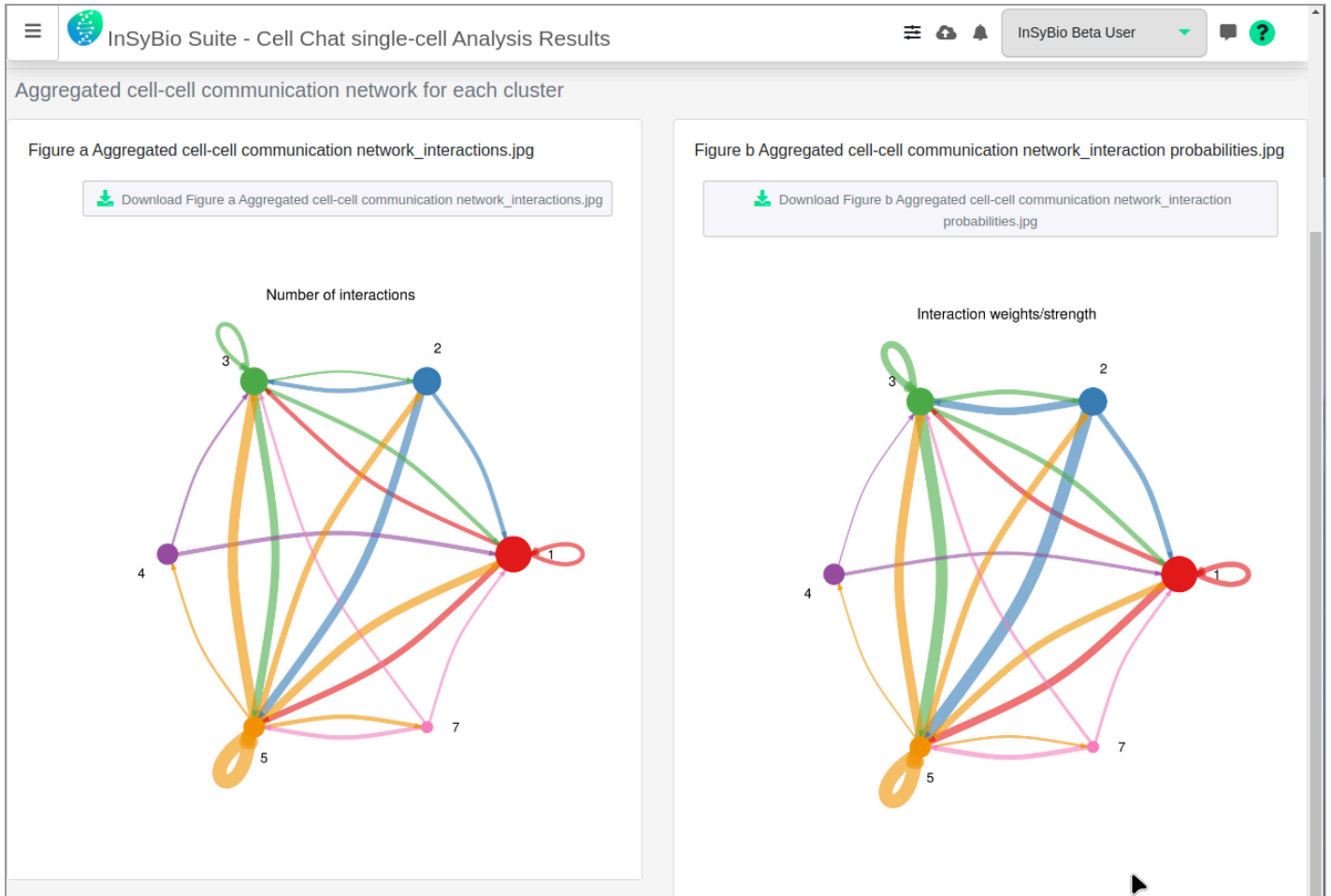
↓ Download Figure Aggregated cell-cell communication network_for cluster_0.jpg

Figure Aggregated cell-cell communication network_for cluster_1.jpg

↓ Download Figure Aggregated cell-cell communication network_for cluster_1.jpg

InSyBio - Intelligent Systems Biology - www.insybio.com - info@insybio.com

Page 80 of 97



☰ InSyBio Suite - Cell Chat single-cell Analysis Results
InSyBio Beta User

	Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
Dashboard	COMPLETED		9	Jan 23, 2024, 8:38:31 AM	00 hours, 11 minutes, 28 seconds	i

Summary Figure for interactions Scatterplot signalling role overall Incoming pattern plots Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities Aggregated cell-cell communication network for cluster for each cluster Chord plots pathway

Circular plots pathway Heatmap plots pathway Grouped chord per pathway plots Intercellular communication network

Distribution of contribution for each ligand-receptor Plots Network signalling role of pathway Plots Scatterplot network signalling plots per pathway All Results Download

Chord Plots Pathway

Figure_chord_plot_pathway_CD99.jpg

[Download Figure_chord_plot_pathway_CD99.jpg](#)

CD99 signaling pathway network

Figure_chord_plot_pathway_CypA.jpg

[Download Figure_chord_plot_pathway_CypA.jpg](#)

CypA signaling pathway network

☰

InSyBio Suite - Cell Chat single-cell Analysis Results

☰
🔔
👤 InSyBio Beta User
?

	Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
← Dashboard	COMPLETED		9	Jan 23, 2024, 8:38:31 AM	00 hours, 11 minutes, 28 seconds	i

Summary Figure for interactions Scatterplot signalling role overall Incoming pattern plots Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities Aggregated cell-cell communication network for cluster for each cluster Chord plots pathway

Circular plots pathway

Heatmap plots pathway Grouped chord per pathway plots Intercellular communication network

Distribution of contribution for each ligand-receptor Plots Network signalling role of pathway Plots Scatterplot network signalling plots per pathway All Results Download

Circular Plots Pathway

Figure_circle_plot_pathway_CD99.jpg

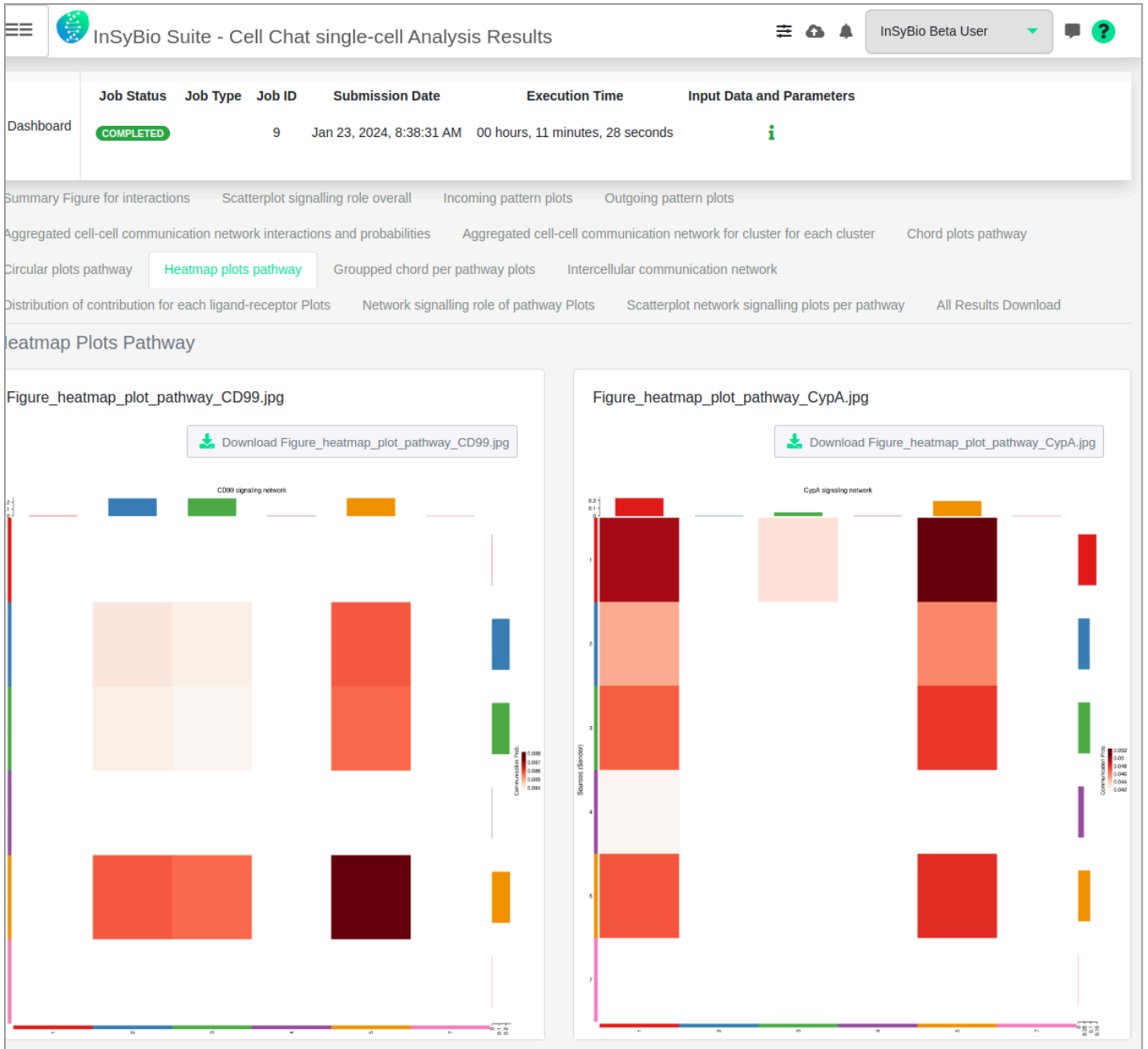
↓ Download Figure_circle_plot_pathway_CD99.jpg

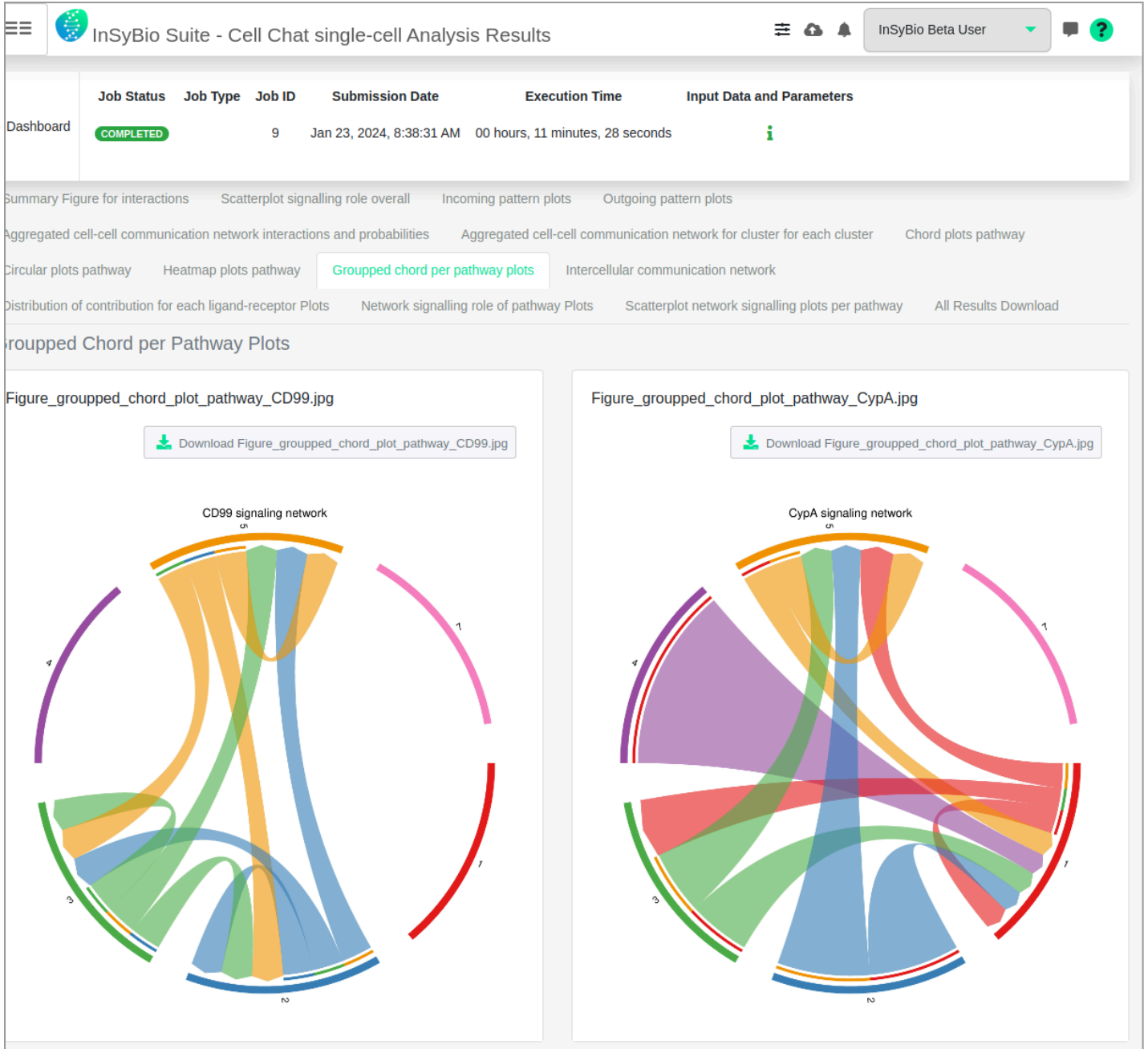
Figure_circle_plot_pathway_CypA.jpg


↓ Download Figure_circle_plot_pathway_CypA.jpg

InSyBio - Intelligent Systems Biology - www.insybio.com - info@insybio.com

Page 83 of 97





 InSyBio Suite - Cell Chat single-cell Analysis Results InSyBio Beta User

[Dashboard](#) **COMPLETED** 9 Jan 23, 2024, 8:38:31 AM 00 hours, 11 minutes, 28 seconds

Summary Figure for interactions Scatterplot signalling role overall Incoming pattern plots Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities Aggregated cell-cell communication network for cluster for each cluster Chord plots pathway

Circular plots pathway Heatmap plots pathway Grouped chord per pathway plots **Intercellular communication network** Distribution of contribution for each ligand-receptor Plots

Network signalling role of pathway Plots Scatterplot network signalling plots per pathway All Results Download

Supplementary CSV files

Supplementary Table 1 intercellular communication network [Download Supplementary Table 1 CSV](#)

Supplementary Table 2 significant pathways [Download Supplementary Table 2 CSV](#)

InSyBio Suite - Cell Chat single-cell Analysis Results

Circular plots pathway Heatmap plots pathway Grouped chord per pathway plots Intercellular communication network

Distribution of contribution for each ligand-receptor Plots Network signalling role of pathway Plots Scatterplot network signalling plots per pathway All Results Download

Distribution of contribution for each ligand-receptor pair to pathway plots for each pathway

Distribution_of_contribution for each ligand-receptor pair_to_pathway_CD99.jpg

Download Distribution_of_contribution for each ligand-receptor pair_to_pathway_CD99.jpg

Contribution of each L-R pair

CD99 - CD99

Relative contribution

Distribution_of_contribution for each ligand-receptor pair_to_pathway_CypA.jpg

Download Distribution_of_contribution for each ligand-receptor pair_to_pathway_CypA.jpg

Contribution of each L-R pair

PPIA - BSG

Relative contribution

InSyBio Suite - Cell Chat single-cell Analysis Results

InSyBio Beta User

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

COMPLETED
9
Jan 23, 2024, 8:38:31 AM
00 hours, 11 minutes, 28 seconds
i

Summary Figure for interactions
Scatterplot signalling role overall
Incoming pattern plots
Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities
Aggregated cell-cell communication network for cluster for each cluster
Chord plots pathway

Circular plots pathway
Heatmap plots pathway
Grouped chord per pathway plots
Intercellular communication network
Distribution of contribution for each ligand-receptor Plots

Network signalling role of pathway Plots
Scatterplot network signalling plots per pathway
All Results Download

Network signalling role of _pathway_CD99.jpg

Download Network signalling role of _pathway_CD99.jpg

CD99 signaling pathway network

Role	1	2	3	4	5	6
Sender	0.1	0.9	0.1	0.1	0.9	0.1
Receiver	0.1	0.1	0.1	0.1	0.1	0.1
Mediator	0.1	0.1	0.1	0.1	0.1	0.1
Influencer	0.1	0.1	0.1	0.1	0.1	0.1

Network signalling role of _pathway_CypA.jpg

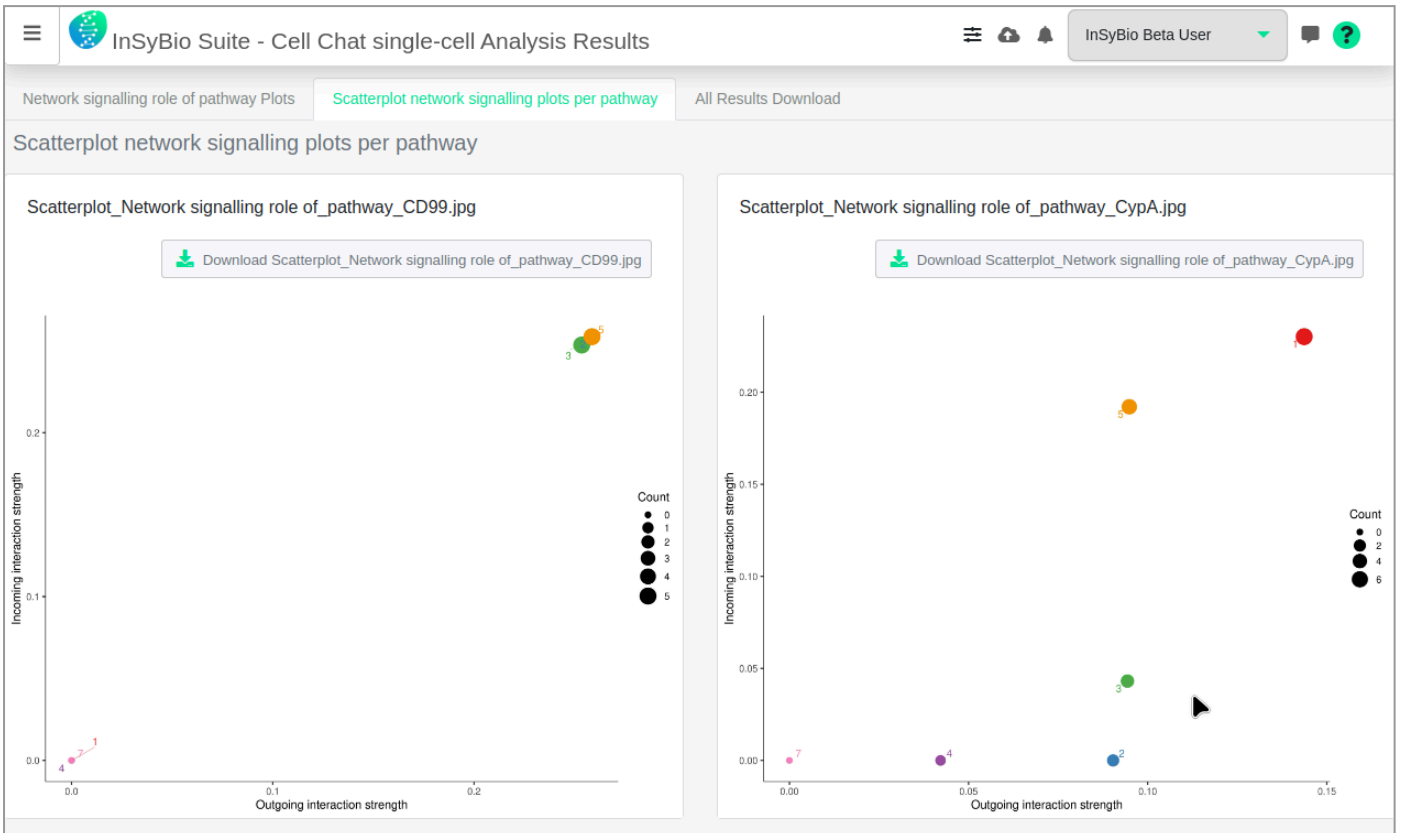
Download Network signalling role of _pathway_CypA.jpg

CypA signaling pathway network

Role	1	2	3	4	5	6
Sender	0.9	0.7	0.5	0.3	0.1	0.1
Receiver	0.1	0.1	0.1	0.1	0.1	0.7
Mediator	0.1	0.1	0.1	0.1	0.1	0.5
Influencer	0.9	0.7	0.5	0.3	0.1	0.1

InSyBio - Intelligent Systems Biology - www.insybio.com - info@insybio.com

Page 88 of 97



InSyBio Suite - Cell Chat single-cell Analysis Results

COMPLETED 9 Jan 23, 2024, 8:38:31 AM 00 hours, 11 minutes, 28 seconds

Summary Figure for interactions Scatterplot signalling role overall Incoming pattern plots Outgoing pattern plots

Aggregated cell-cell communication network interactions and probabilities Aggregated cell-cell communication network for cluster for each cluster Chord plots pathway

Circular plots pathway Heatmap plots pathway Grouped chord per pathway plots Intercellular communication network Distribution of contribution for each ligand-receptor Plots

Network signalling role of pathway Plots Scatterplot network signalling plots per pathway **All Results Download**

All Results Download

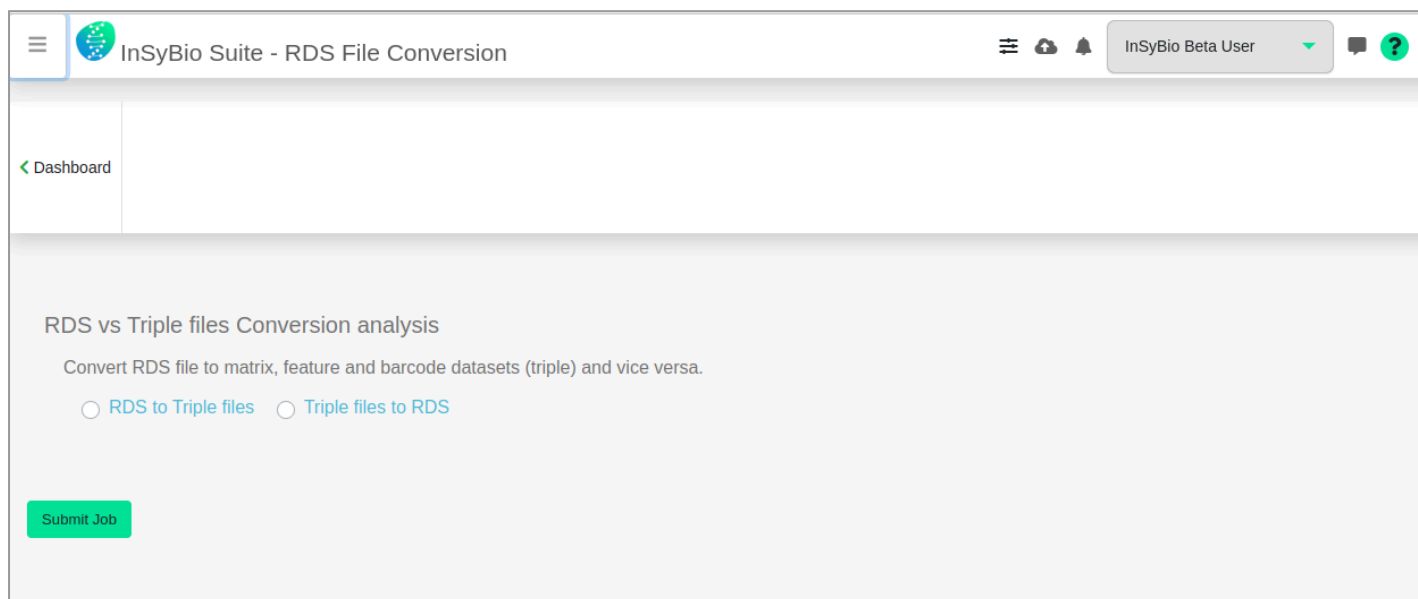
Cell Chat single-cell Analysis Results [Download](#)

Compressed Folder [Folder](#)

RDS File Conversion

You can convert a Seurat object file (.rds format) file to 10X Matrix, Features and Barcodes datasets (triple) and vice versa. Depending on the selected option,

- RDS to triple
- Triple to RDS



To start the RDS File Conversion:

Click in the menu “InSyBio ncRNASeq” → “single-cell RNA-Seq Data Analysis” → “single-cell RNA-Seq Pipeline Dashboard”, select the “Add new job” button and then choose the “RDS file conversion” option. Then depending on the selected option do the following steps:

- RDS to triple:
 - Select or upload a Seurat object and the algorithm will convert it to matrix, features and barcode datasets.

InSyBio Suite - RDS File Conversion

InSyBio Beta User

< Dashboard

RDS vs Triple files Conversion analysis

Convert RDS file to matrix, feature and barcode datasets (triple) and vice versa.

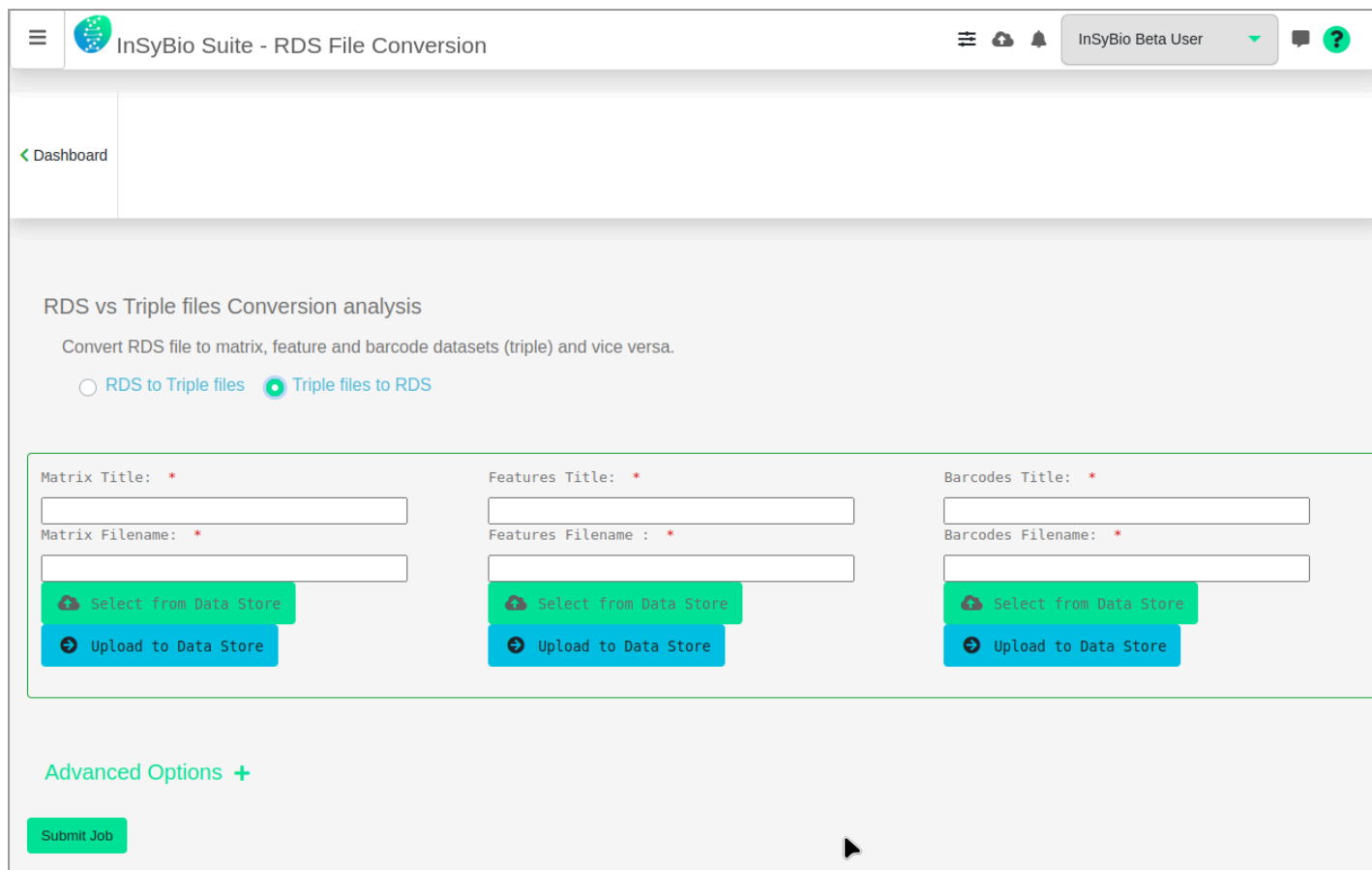
RDS to Triple files Triple files to RDS

RDS File ⓘ

Title:

Filename:

- Triple to RDS:
 - Select or upload the three matrix, features and barcodes files and the algorithm will convert it to a Seurat object file.



InSyBio Suite - RDS File Conversion

InSyBio Beta User

< Dashboard

RDS vs Triple files Conversion analysis

Convert RDS file to matrix, feature and barcode datasets (triple) and vice versa.

RDS to Triple files
 Triple files to RDS

Matrix Title: *	Features Title: *	Barcodes Title: *
<input type="text"/>	<input type="text"/>	<input type="text"/>
Matrix Filename: *	Features Filename: *	Barcodes Filename: *
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="Select from Data Store"/>	<input type="button" value="Select from Data Store"/>	<input type="button" value="Select from Data Store"/>
<input type="button" value="Upload to Data Store"/>	<input type="button" value="Upload to Data Store"/>	<input type="button" value="Upload to Data Store"/>

[Advanced Options +](#)

- Select if you want to manually configure other parameters of the job. If you don't, our Default Options will be applied. Possible manual options are:
 - First filtering:
 - Minimum cells
 - Minimum features
 - Secondary filtering:
 - nFeature_RNA with lower and upper limits
 - nCount_RNA with lower and upper limits
 - Feature Extraction Method
 - Shared Nearest Neighbor (SNN) Graph
 - K parameter (k-nearest- neighbor)
 - Clustering
 - Resolution parameter

Advanced Options +

Cluster annotation

Species:

Tissue  :

First filtering

Minimum cells:

Minimum features:

Secondary filtering

nFeature_RNA ? :

Lower limit:

Upper limit:

nCount_RNA ? :

Feature Extraction Method

Shared Nearest Neighbor (SNN) Graph

k parameter (k-nearest-neighbor):

Clustering

Resolution parameter ? :

- Submit your job pressing the respective button.

To view the results:

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

InSyBio Suite - Single Cell RNA-Seq Differential Expression Pipeline Dashboard

InSyBio Beta User

Filter Jobs Show All 13 Completed 1 Running 0 Pending 4 Error

Status	Job ID	Job Type	Input File(s)	Submission Date	Start Execution Date	Completion Date	Current Step	Actions
Completed	21	RNASeq Single Cell Velocity Analysis		11/26/73, 3:22 AM	1/16/24, 1:59 PM	-	Secondary Single Cell Analysis	View Results
Completed	20	Deconvolve Data against single-cell RNA-seq Analysis		8/12/11, 6:41 AM	1/15/24, 2:14 PM	1/15/24, 2:15 PM		View Results
Completed	19	RNASeq Single Cell Velocity Analysis		3/29/80, 9:11 PM	1/12/24, 8:51 AM	1/12/24, 9:06 AM	Secondary Single Cell Analysis	View Results
Completed	18	Cell Chat Analysis		12/19/68, 6:13 AM	1/12/24, 8:29 AM	1/12/24, 8:33 AM	Single Cell Alignment	View Results
Completed	17	RDS Conversion		2/9/59, 5:58 AM	1/11/24, 11:16 AM	1/11/24, 11:18 AM	Secondary Single Cell Analysis	View Results
Completed	15	RDS Conversion		9/21/76, 5:35 AM	1/17/24, 10:54 AM	1/17/24, 10:55 AM	Secondary Single Cell Analysis	View Results

After the analysis, you can select the View Results in the Actions column and view the produced files, that are separated according to the step that they were produced.

- RDS to triple: The 10X triple files, matrix, barcodes and features files are produced and ready to be downloaded from the Results Files tab.

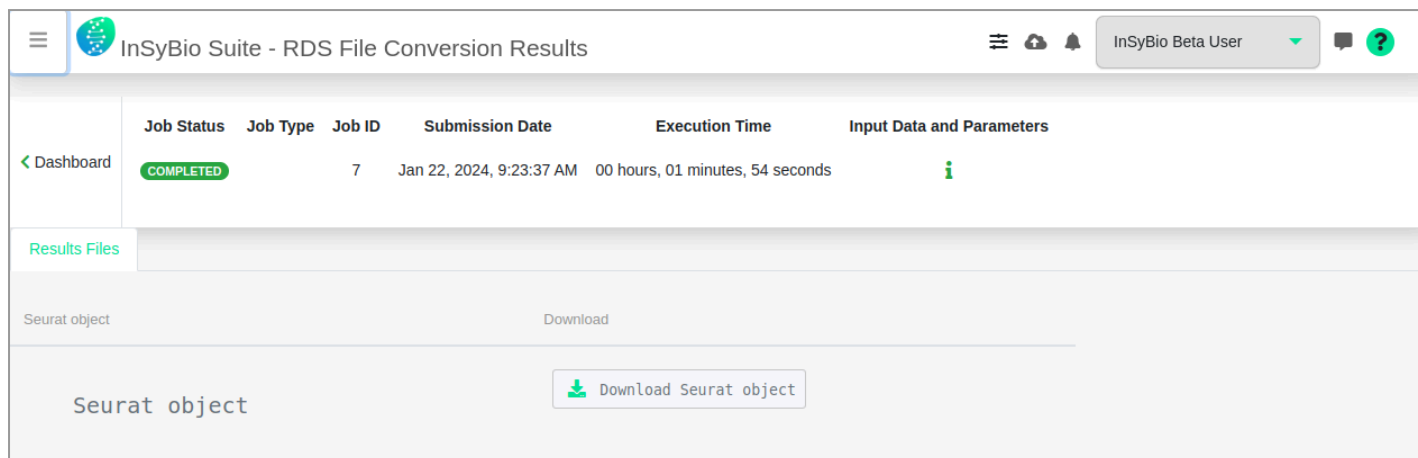
The screenshot displays the InSyBio Suite interface for RDS File Conversion Results. The top navigation bar includes a menu icon, the InSyBio logo, the page title 'InSyBio Suite - RDS File Conversion Results', and user information 'InSyBio Beta User'. Below the navigation bar, a table lists job details:

Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
COMPLETED		8	Jan 23, 2024, 8:27:10 AM	00 hours, 03 minutes, 11 seconds	i

The 'Results Files' tab is selected, showing a 'Fastq Dataset' with a 'Download' button. Below this, three file download options are listed:

- Features File: Download Features File
- Matrix File: Download Matrix File
- Barcodes File: Download Barcodes File

- Triple to RDS: The produced Seurat object can be downloaded from the Results Files tab.



InSyBio Suite - RDS File Conversion Results

InSyBio Beta User

Job Status	Job Type	Job ID	Submission Date	Execution Time	Input Data and Parameters
COMPLETED		7	Jan 22, 2024, 9:23:37 AM	00 hours, 01 minutes, 54 seconds	i

Results Files

Seurat object

Download

Seurat object

Download Seurat object

How to get InSyBio ncRNASeq

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

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

About Us

InSyBio Ltd 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 InSyBio Inc. 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 2024 InSyBio Inc Reproduction without written permission is completely forbidden.