



InSyBio

Intelligent Systems Biology

User Manual

Analyze non-coding RNAs with InSyBio ncRNASeq

November 2019

Insybio Suite v2.6

www.insybio.com

Introduction

ncRNASeq is a non-coding RNA analysis tool for the prediction and analysis of:

- non-coding RNAs
- miRNA target genes

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 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 own datasets.

With InSyBio ncRNASeq you can:

- a) Calculate 58 RNA genes-related features
- b) Predict miRNAs
- c) Calculate 124 miRNA target sites 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 pipeline on your RNASeq data and perform Differential Expression Analysis

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.

The screenshot shows the InSyBio ncRNASeq web interface. The sidebar on the left contains navigation options: InSyBio Interact, InSyBio ncRNASeq, non-coding RNA Analytics, ncRNA Feature Calculation (selected), miRNA Prediction, miRNA Target site Feature Calculation, miRNA Target site Prediction, miRNA Target Prediction, ncRNASeq Knowledge Base, miRNA and transcript search, and RNA-Seq Data Analysis. The main panel displays the 'ncRNA Feature Calculation' module with a 'Sequences' field containing '75 sequences including pre-miRNAs, random cds and snoRNAs' and a 'File Title' field containing 'dsfile1444763196_4314.txt'. Below these fields are two buttons: 'Select file from Data Store' and 'Go to Data Store to Upload File'. A 'Start calculation' button is located on the right. Below the input fields is a table showing the execution status of various jobs.

Status	Process ID	Information	Submission Date	Start Execution Date	Completion Date	Actions
Completed	11	test	11/30/18 9:51 AM	11/30/18 9:51 AM	11/30/18 9:51 AM	View Results
Completed	9	test	11/15/18 8:59 PM	11/15/18 8:59 PM	11/15/18 8:59 PM	View Results
Completed	8	sequences75_premiRNAs_cds_snoRNAs2222	11/8/18 2:35 PM	11/8/18 2:35 PM	11/8/18 2:35 PM	View Results
Completed	7	75 sequences including pre-miRNAs, random cds and snoRNAs	11/8/18 8:48 AM	11/8/18 8:49 AM	11/8/18 8:49 AM	View Results
Completed	6	test	11/7/18 12:04 PM	11/7/18 12:04 PM	11/7/18 12:04 PM	View Results
Completed	3	75 sequences including pre-miRNAs, random cds and snoRNAs	9/26/18 10:42 AM	9/26/18 10:42 AM	9/26/18 10:42 AM	View Results
Completed	1	75 sequences including pre-miRNAs, random cds and snoRNAs	8/17/18 7:04 AM	8/17/18 7:06 AM	8/17/18 7:08 AM	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.

The results are presented in 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 e 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	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,	8/17/18 7:11	8/17/18 7:11 AM	8/17/18 7:11	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.
- 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 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 at the Actions column and view the calculated features.

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

COMPLETED 2 Aug 17, 2018 7:11:08 AM 00 hours, 00 minutes, 02 seconds [Export Results](#)

Sequence	Prediction Score	Prediction	G+C	A+U	AA	AC
> hsa-mir-26a-1 MI000083 GUGGCCUCGUUCAAUAUCCAGGAUAGGCUUGCAGGUCUCCAAUUGGGCUAUUCUUGGUUACUUGCACGGGGACGC	1.02096	miRNA	55.844	44.156	3.947	3.947
> random_sequence_from_cds_1 GAGGGCAGGGGGCAGUCCAAUCUCCAGGCUUAGUAGUCCAGGGGCUUGGUGCCGCCGCGCAGCGGCAGACUGUCCUGUGGGCCGUGCACA	-0.893914	pseudomiRNA	69.072	30.928	1.042	4.167
> snoRNA_1 AAAGUGAGUGAUGAAUUGUUGGCAUUAUGAAUUAUUUUUUGAUUAAACCCUAAACUCUGAAUGCC	NaN	other	32.857	67.143	14.493	2.899
> hsa-mir-32 MI000090 GGAGAUUUUCCACAUUUAUUGUUGCAUGUUGUACGGCCUCAUUGCAUUUUGUUGUGUGAUUUUUUC	1.06056	miRNA	38.571	61.429	4.348	4.348
> hsa-mir-199a-1 MI0000242 GCCAACCCAGUGUUCAGACUACUCCUAGGAGGCGUCUCAAUGUGUACAGUAGUCGACAUUGGUUAGGC	0.92389	miRNA	50.784	49.296	2.857	7.143
> hsa-mir-148a MI0000253 GAGGCAAGUUCUGAGACUCCGACUCUGAUUGAUGAAGUCAGUCACUACAGAAUUUGUCUC	1.17143	miRNA	45.588	54.412	5.97	8.955

Showing 1 to 25 of 75 entries

The results are presented in 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 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.

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: mmas462, miRNAs: mimas462	11/8/18 1:45 PM	11/8/18 1:45 PM	11/8/18 5:51 PM	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: 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_S_0_shuffled_targets, miRNAs: genes_5_S_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 at the Actions column and view the calculated features.

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 UACAGUACUGUAUACUGAA	> 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 UACAGUACUGUAUACUGAA	> NM_004456EZH220478051 Homo sapiens TTCAGGAACCTCGACTGTG	8	6	14	3	3	6	5	3	8	0	16	16	2	2	4	-2	
> [hsa-miR-101] Homo sapiens UACAGUACUGUAUACUGAA	> NM_181833NF217220301 Homo sapiens TACAAGAGATTCTCTGCTCA	4	3	7	2	2	4	2	1	3	8	22	30	0	0	0	8	
> [hsa-miR-101] Homo sapiens UACAGUACUGUAUACUGAA	> 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 esntane	> NM_001039111TRIM7117890240 Hmn esntane	9	3	12	3	2	5	6	1	7	-2	25	23	0	0	0	-2	

The results are presented in 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)	casl1s, casl2s, casl3s, casl4s, casl5s, casl6s, casl7s, casl8s	structural
number of asymmetric loops of length 1-7 and greater than 7 in out-seed part (8 features)	casl1os, casl2os, casl3os, casl4os, casl5os, casl6os, casl7os, casl8os	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	thermodynamic
free energy of the out-seed part	mfeos	thermodynamic
free energy of the total miRNA-mRNA alignment structure	mfe	thermodynamic
free energy of the target sequence	mfet	thermodynamic
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	thermodynamic
difference in the free energies of the total miRNA-perfect target alignment structure and the total miRNA-mRNA alignment structure	dmfe	thermodynamic
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	s106	positional

seed positions.		
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.

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: 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: mnas462, miRNAs: mirnas462	11/8/18 1:45 PM	11/8/18 1:45 PM	11/8/18 5:51 PM	View Results

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: targethsa-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 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 at the Actions column and view the predictions and calculated features.

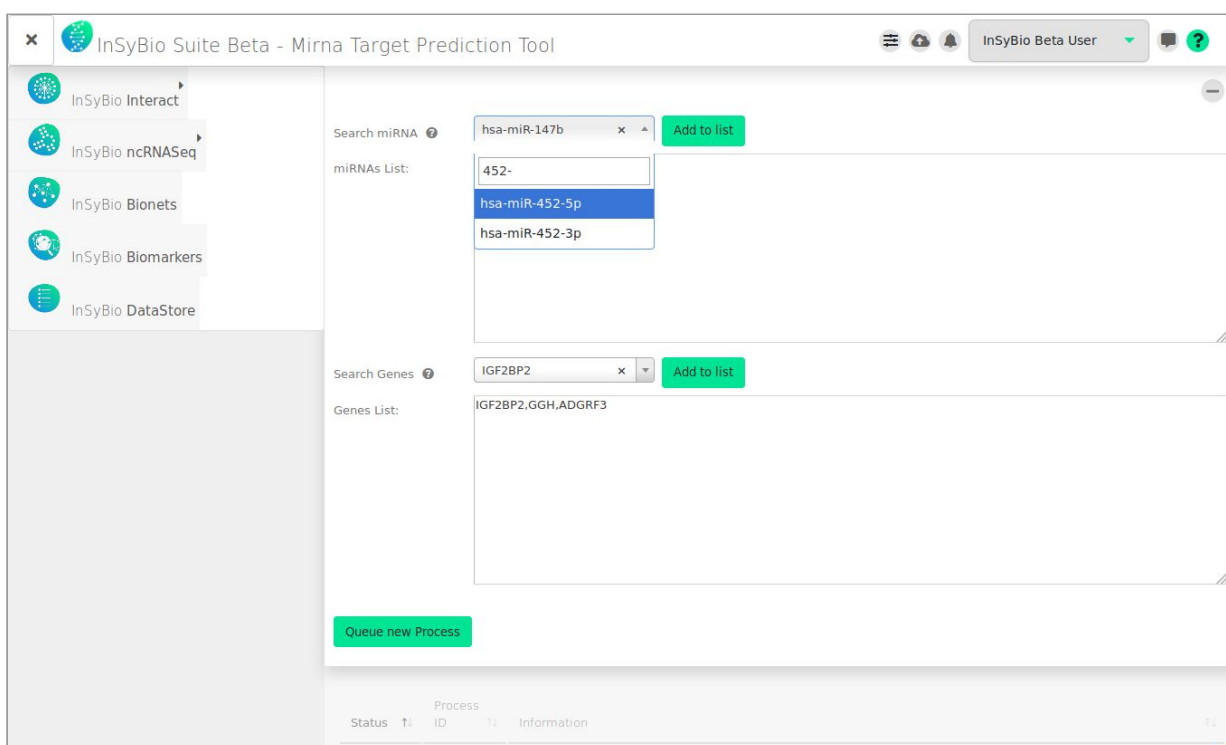
miRNA Sequence	Target Sequence	Prediction Score	Prediction	mats	matos	mat	gcmats	gcmatos	gcmat	aumats	aumatos	aumat	unps	unpos	unp
> [hsa-miR-101] Homo sapiens UACAGUACUGUGUAUCUGAA	> NM_004456EH220478051 Homo sapiens TGAATTTGCAAAGTACTGTA	0.963256	Target	9	2	11	3	1	4	6	1	7	-2	22	20
> [hsa-miR-101] Homo sapiens UACAGUACUGUGUAUCUGAA	> NM_004456EH220478051 Homo sapiens TTCAGGAACCTCGAGTACTGTG	1.2725	Target	8	6	14	3	3	6	5	3	8	0	16	16
> [hsa-miR-101] Homo sapiens UACAGUACUGUGUAUCUGAA	> 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 UACAGUACUGUGUAUCUGAA	> NM_001039111TRIM7117890240 Homo sapiens ACAACATTGCTTAAGTCTACCTCA	-0.880751	no Target	1	5	6	0	2	2	1	3	4	14	21	35

The results are presented in 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 the 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 in order 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 in order to calculate a prediction score.



To start the prediction:

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

- Select 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-5p,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,SMOC1,FGF2,MMRN2,PRSS3,VEGFB,ADAM21,ADAMTSL4,C10TNF4,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, A18G-A51, 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, there you can view the status of your current and previous miRNA Target Predictions. At completion of the calculation you can select the View Results at the Actions column and view the results.

ite Beta - 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 in your screen in a browse-able 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.

Bio Suite Beta - 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 AGGCCTCATGAATGCAGCAAATGTGGAAGCGCCTCAACTCAAGATCTATCATCATTTAGCTCCTGAAAGTCCACACTTA 80 81 AGTAGAGCCTTAGACCTACAGGGAAGTGTCTCTGTAGTATTGTAGCAGTAGAGAGCCTTTGTGAGGGAGCCATCTG 160 161 CCTGAAGTTGAACCTCATTCTTCCTTGTCTCTGGTAGAAACCATCTACCCCTACCACTTGCACAGTGGGCACTGGT 240 241 CACTCCTATGTGCTAAGCAAGGCAGACATCTGTGTCTCTTAAGTCTTTGGAGGAATCTTGAGCAGTCAAGCCTT 320 321 TAGASAAAATTCATTTCTTTCTGACTGATCACASCATACGTGTGACCCAGTTTGGGTGAGGAGGCCAGCCCTGGTT 400 401 CTGCTGGACACTTATGTGCAAGGATTCCTTCATGTAATTTCTTGGTCTCACAAGACACTTGGTCACTTCTCCAGCCCTC 480 481 ATGTCACCACCTGGTGAATGGCTGCCTCACATTGCTCCAGTTTGTGCACTAATAAAAAGCCTTATATTTGAATCTACCTGT 560 561 AGTCTTGGGGTCTGTTTACTGTGTGGGGTGCCTGGGAGACAGACTTCAACTATATGAAGGAATGGATGGCTTTTGTG 640 641 GGCCCTCGCAGGAAAGTAAAGTACAGAGTAATTTCTAATTTCTGGTTTGGTCACTACTGCTTTGCTACCTAAAATCTCCT 720 721 AGGAAAAATGCAAGGTTTGGTTATTTCTAATTTGTGGCCTGGATCCCTATTTCTTGTGAGACTAGAGGTCATCCCTGA 800 801 GGAGAGGCCAGCTGTTATGACAAGCATGTGCTTCAGGGAATAGGACAATTTATTCATTGTTCCAGAGGATGTGAT 880 881 ATGATGCCAGTGTGCTGAGAAGCTTTTCATGGGTTCTATAAGGAGCATGCCCTGATATCAAACATTCATAGGCCG 960 961 ATGTCACGCAGAAACACCGGAGTCACATGTGAAGTGAATTTGGTACAGAAATACCTGGGTATTTCTGTACTGTGTGTA 1040 1041CTGTAGCAAACTAGTTGGAAATGCTCTTATAAAGTACATTTACAATCTTCCCGTACTGTGGCTTGGAGCAGTCAT 1120 1121AGGACCTAGAAATCTGTGTATGTCCTAATAGCTGAGGTTATTTTCAGCAAAAATAATTAAGGGTTTTATTTTTTAATCT 1200 1201TGTGGTTTTCTAGGTTGTTACCTCAAGTGCATTTGCTGTAGAGGAGCAAAAAGGAGGATAAAGATAACAGAAAGCCTAT 1280 1281AGGCCAGGGATGATTGATAGCTCTTGTGATTTCCACAGTGTGCTGTGTCTCAAATGCCCACAGCCTTCATTGCTTG 1360 1361CAACATTTCTGATGAGGAGCTACTGTTGCTTCCCGAGCCCTGAAGAGAGAGTGCAGTCAACATGAGATGCTA 1440 1441GGCATTCTGGTTCTGAAAGTGGGTGATCAGATACTTTATGTGAAACATGTTTACAACCTTTCTGATGTGTAAGT 1520 1521ACATGGCCATAGTTTACATCCATTTATGGTGTATAATTTGAAGAGTTTGTCTACAAGCCTGTGAAACCATAAATCATGATC 1600 1601ATGAACATATTGATGATTCACCTCTGCTGTTTTACAATCTCTGCTGTACTTTCCAGGCCCTCAGGAGTCCCTGCTATT 1680 1681ACTTTCCCTACAGAGAAATGTTTGTGTTTTCTAGGATTTATGTGAATTTGAACGTAATAATCTACTCCATTTTCTCT 1760 </pre>

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

Score : 1.5224538611539185
 TBACACTTGTCTATTTCCAGCCTCCATG
 ||||| |||
 -----AGAGGCCGCGGAAGUG

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

Accession	Name	Sequence	FASTA	Evidence	Experiment
MIMAT000256	hsa-miR-181a-5p	39 aacaucaacgcugucgugagu 61	Download	Experimental	cloned [2,4-6]
MIMAT0004558	hsa-miR-181a-2-3p	77 accacugaccguugacugacuacc 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 similarity of the related stem-loop to another stem-loop or found in 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, J, 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

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 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 shows the InSyBio ncRNASeq interface. At the top, a search bar contains 'zik1' and a 'Show results' button. Below the search bar, there are two tabs: 'Transcripts' (selected) and 'Genes'. The 'Transcripts' tab displays a list of transcripts with columns for 'Ensemble Transcript id' and 'Transcript name'. The first transcript, ENST00000307468, is highlighted. To the right of the list, the details for 'Transcript: ZIK1-004 ENST00000307468' are shown. The 'Information' tab is active, displaying various fields: 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), Transcript 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), and Description (zinc finger protein interacting with K protein 1 [Source:HGNC Symbol;Acc:HGNC:33104 External Link to HGNC]). At the bottom of the details, there are buttons for '3'UTR Visualization' and 'Download'.

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>GAGTGTACAGTCAAAGGCAGGTTTCATCCACACAGAAGACTCAATCCTGTGAGATGTGTGCCAGTCTGAAAGATATT TTGCATCTAGCTGATCTCCCTGGGCAGAAACCATCTTGGTTGGAGAATGTACAAACCATCACCAGCACCAGAAGCATCA CAGTGCAGAAAGAAATCCTTGAAGAGGGACATGGACAGAGCCATATGTGAAGTGCTGCATTTCTGTATGTCTTGAAGC CCTTCGAAATGGGAGGTTGGAAAGGACCTCCAGCCATGTTGCGGCTTCTGAGGTCCTGGTCTTTCTGGAGGCAAG AAACCCGGCACAATTACTGAATGTGGGAGGACATTCAGTCAAAGGCTCATTACAAGTCAGGTGAATGTGGGAAGGC TTCCAGGCACAAACACACTCCTGTTTACCATCCAAGAGTCTACACTGGAAAAAGCTTTATGAGTGTAGCAAATGTGGGA AAGCCTTCGTGGCAAGTACTCACTTGTTCAGCACAGAGAGTCCATACTGGAGAAAGGCCCTGGGAGTGAATGAATGT GGAAATTTCTTAGCAAACCTCCACCTGAATGATCATCGGAGAATCCACCCGGAGAAAGGCCCTTATGAGTGCAGCGA ATGTGGAAAAATTTTAGCAAAACTCCAGCCTTGTGACCACCAGAAAAATACACTGGAGCAAGGCCCTTATGAGTGTA GCCAGTGTGGGAAATCCTTTAGCCAAAAGCCACCTTGTAAACACCAAAGAGTTCACACTGGAGAAAGGCCCTATAAG TGTGGTGAATGTGGGAATTCCTTAGTCAAAGTGCATTCTTAATCAACACCGAAGAATTCACACTGGAGCAAGGCCCTA TGAGTGTGGCCAGTGTGGGAAATCCTTAGTCAAAGGCTACCCTCATTAAACACAGAGAGTTCACACTGGAGAAAGGC CTTATAAGTGTGGTACTGTGGGAAATCCTTAGTCAAAGCTCCATCCTATTCAACACCGGAGAATTCATACTGGAGCA AGGCCCTTATGAGTGTGGCAGTGTGGAAAGTCCCTTAGCCAAAAGCTGGTCTCATTCAACACCAAGTGGTTCACACTGG AGAAAGGCCCTTATGAGTGAACAAATGTGGGAATTCCTTAGCCAAATGCTCCAGCCCTACATACACCAAAAATGTGATA ACACATAGAGGCCCTCATGAATGCAGCAAATGTGGAAGCGCCTTCAACTCAAGATCTATCATCATTAGTCTCTGAAAGTC CACACTTAAGTAGAGCCTTAGACCTACAGGGAAGTGTCTCTGTAGTATTGTAGCAGTAGAGAGCCTTTGTGAGGGGA GCCATCTGCTGAAGTTGAACCTCATTCTTCTTGTCTCTGGTAGAAACCATCTACCCTTACCACCTGGCACAGTGG GCACCTGGTCACTCTATGTCTAAGACAAGGCAGACATCTGTGTCTCTTAAAGTCTTTGGAGGAAATCTTGAGCAGTC TAAGCCTTTAGAGAAAATTCATTCTTTTTCTGACTGATCACAGCATACGTGTGACCAAGTTTGGGTGAGGAGGCCAG CCTTGGTTCTGCTGGACACTTATGTGAAGGATTCCTTCAATGAAATTCCTGGTCTCACATGACACTTGGTCACTTCTTC CAGCCTCCATGTACCACGTGGTGAATGGCTGCCTCACATTGCTCCAGTTTGTGCACTAATAAAGCCTTATATTTGAAT CTACCTGATGCTTGGGTTCTGTTTACTGTGTGGGTTGGCTGGGAGACAGACTTCAACTCTATATGAAGGAATGGATGG CTTTGTGGGCTCTGCAGGAAAGTAAGATGACAGAGTAATCTAATCTGGTTTTGGTCATACCTTGGCTTGTACCTAA AATCCTAGGAAAAATGCAAGGTTTTGGTTATTCTAATTTGGGCTGGATCCCTATTTCTGTGAGACTAGAGGT CATCTGAGGAGAGGCCAGCTGTTATGACAAGCATGTGTGCTTCAAGGAAATAGGACAATTTTATCCATTGTTTCCAGAG CATCTGATGATGAGGAGGCTGCTGAGAGGCTTTTATGGGTTCTATAGGAGGCTCCCTGATGATGAAATCTTC</p>
ENST00000456074	ZIK1P1-001	
ENST00000536878	ZIK1-002	
ENST00000597219	ZIK1-006	
ENST00000597850	ZIK1-001	
ENST00000598689	ZIK1-007	
ENST00000598726	ZIK1-008	
ENST00000599456	ZIK1-003	
ENST00000600000	ZIK1-005	

Genes information

The screenshot shows the 'Gene Search Tool' interface. A search box contains 'zik1' and a 'Show results' button is visible. The results are displayed under the 'Genes' tab. A table lists two genes: ZIK1 (ENSG00000171649) and ZIK1P1 (ENSG00000237426). The 'Information' tab for ZIK1 is expanded, showing details such as Name - Source (ZIK1 (HGNC Symbol)), Description (zinc finger protein interacting with K protein 1), Location (Chromosome 19: 57578456-57593777 forward strand), Transcript count (8), Biotype (protein_coding), Status (Known), Annotation method (Automatic from Ensembl and Havana manual curation), and Version (ENSG00000171649.11).

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 to links to their information.

This screenshot shows the 'Transcript Table' for the gene ZIK1. The 'Transcript Table' tab is selected, displaying a table with 8 rows of transcript information. The table columns are '#', 'Ensemble id', and 'Name'. The transcripts listed are ZIK1-002, ZIK1-006, ZIK1-001, ZIK1-007, ZIK1-008, ZIK1-003, ZIK1-005, and ZIK1-004.

#	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 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 sequences quality, then trimm 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 of FeatureCounts, that creates text files with the quantity of each gene. At 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 either check 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.

InSyBio Suite Beta - RNA-Seq Differential Expression Pipeline

RNA-Seq Data: Paired-end Single-ended

Condition Control: Normal

Title Read 1: HBR rep1 read1 Title Read 2: HBR rep1 read2
Filename Read 1: dsfile1557128487_9359.gz Filename Read 2: dsfile1557128516_9128.gz
Select from Data Store Upload to Data Store Select from Data Store Upload to Data Store

Title Read 1: HBR rep2 read1 Title Read 2: HBR rep2 read2
Filename Read 1: dsfile1557128550_6204.gz Filename Read 2: dsfile1557128587_1781.gz
Select from Data Store Upload to Data Store Select from Data Store Upload to Data Store Delete Pair Add Pair

Condition 1: Turner
Title Read 1: UHR rep1 read1 Title Read 2: UHR rep1 read2
Filename Read 1: dsfile1557128760_6526.gz Filename Read 2: dsfile1557128859_1587.gz
Select from Data Store Upload to Data Store Select from Data Store Upload to Data Store Add Pair

Condition 2: Treated
Title Read 1: UHR rep3 read1 Title Read 2: UHR rep3 read2
Filename Read 1: dsfile1557129113_6485.gz Filename Read 2: dsfile1557129149_2179.gz
Select from Data Store Upload to Data Store Select from Data Store Upload to Data Store Delete Pair Add Pair

Add Condition Clear All Files

InSyBio Suite Beta - RNA-Seq Differential Expression Pipeline

RNA-Seq Paired-end Single-ended

Data: *

Condition Control: Control

Title: MDA-MB-231_1.fq.gz

Filename: dsfile1564676913_4027.gz

Select from Data Store Upload to Data Store

Title: MDA-MB-453_1.fq.gz

Filename: dsfile1564671163_2990.gz Delete File

Select from Data Store Upload to Data Store

Add File

Condition 1: Treated

Title: HCC1937_1.fq.gz

Filename: dsfile1564736477_8378.gz

Select from Data Store Upload to Data Store

Title: Hs578T_1.fq.gz

Filename: dsfile1564739672_2540.gz Delete File

Select from Data Store Upload to Data Store

Add File

Add Condition

Clear All Files

Options

- Name Conditions/Group of files you want to Differentially Express.
- 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 to 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


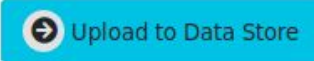
Filename: dsfile1573556494_9916.fa

Select the reference genome (GTF): *

Title: chr22 GTF

Filename: dsfile1573556655_8832.gtf

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	1	RNASeq Analysis	HBR: 1. HBR rep1 read1, HBR rep1 read2, 2. HBR rep2 read1, HBR rep2 read2, 3. HBR rep3 read1, HBR rep3 read2 UHR: 4. UHR rep1 read1, UHR rep1 read2, 5. UHR rep2 read1, UHR rep2 read2, 6. UHR rep3 read1, UHR rep3 read2	5/6/19 7:55 AM	5/6/19 11:53 AM	5/6/19 12:09 PM	Differential Expression Analysis	View Results
Error	2	RNASeq Analysis	Control: 1. HBR rep1 read1, HBR rep1 read2, 2. HBR rep3 read1, HBR rep3 read2 tumor : 3. UHR rep1 read1, UHR rep1 read2, 4. UHR rep2 read1, UHR rep2 read2, 5. UHR rep3 read1, UHR rep3 read2	5/6/19 10:23 AM	5/6/19 12:18 PM	5/6/19 12:21 PM	Differential Expression Analysis	View Details
Completed	3	RNASeq Analysis	hbr_condition: 1. HBR rep1 read1, HBR rep1 read2, 2. HBR rep2 read1, HBR rep2 read2, 3. HBR rep3 read1, HBR rep3 read2 uhr_condition: 4. UHR rep1 read1, UHR rep1 read2, 5. UHR rep2 read1, UHR rep2 read2, 6. UHR rep3 read1, UHR rep3 read2	5/6/19 12:28 PM	5/6/19 12:28 PM	5/6/19 12:34 PM	Differential Expression Analysis	View Results
Completed	5	RNASeq	BC cell lines - Group 1b: 1. MDA-	8/2/19 10:57	8/2/19	8/2/19	Differential	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 they were produced.

InSyBio Suite Beta - RNA-Seq Differential Expression Pipeline Results

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

Deseq2 Reports | Initial FastQC Reports | Trimmed FASTQ Files | Trimmed FastQC Reports | Alignment Files | Read Count Files | Next Actions

HBR_UHR








- 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.













Deseq2 Reports | **Initial FastQC Reports** | Trimmed FASTQ Files | Trimmed FastQC Reports | Alignment Files | Read Count Files | Next Actions

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 Files, BAM Files, and Run Info. Each section has a 'Download' link. The SAM Files 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 Files 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 word '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 file HBR_1.bam (HBR_1.bam);	File
Job-1 BAM file HBR_2.bam (HBR_2.bam);	File
Job-1 BAM file HBR_3.bam (HBR_3.bam);	File
Job-1 BAM file UHR_1.bam (UHR_1.bam);	File
Job-1 BAM file UHR_2.bam (UHR_2.bam);	File
Job-1 BAM file UHR_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.





Deseq2 Reports	Initial FastQC Reports	Trimmed FASTQ Files	Trimmed FastQC Reports	Alignment Files	Read Count Files	Next A
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	
Deseq2 Reports	Alignment Files	Read Count Files	Predicted ncRNAs	Next Actions
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 shows the 'Next Actions' tab in the InSyBio Suite for the HBR_UHR analysis. The interface is organized into two sections: 'Molecule Quantification Files per Condition' and 'Full Molecule Quantification File and Associated Labels'. Each section contains a list of files with 'Download' and 'Next Action' options.

File Name	Download	Next Action
Job-1 MQ file HBR_UHR_diffexpr-MQHBR.csv (HBR_UHR_diffexpr-MQHBR.csv);	 File	--Select Action--
Job-1 MQ file HBR_UHR_diffexpr-MQUHR.csv (HBR_UHR_diffexpr-MQUHR.csv);	 File	--Select Action--
Full Molecule Quantification File and Associated Labels		
Job-1 MQ file HBR_UHR_diffexpr-MQ.csv (HBR_UHR_diffexpr-MQ.csv);	 File	--Select Action--
Job-1 label file HBR_UHR_diffexpr-labels.txt (HBR_UHR_diffexpr-labels.txt);	 File	--Select Action--

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 built a classification model able to predict to which of the two conditions a potential new sample belongs.

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

Copyright 2019 InSyBio Ltd. Reproduction without written permission is completely forbidden.