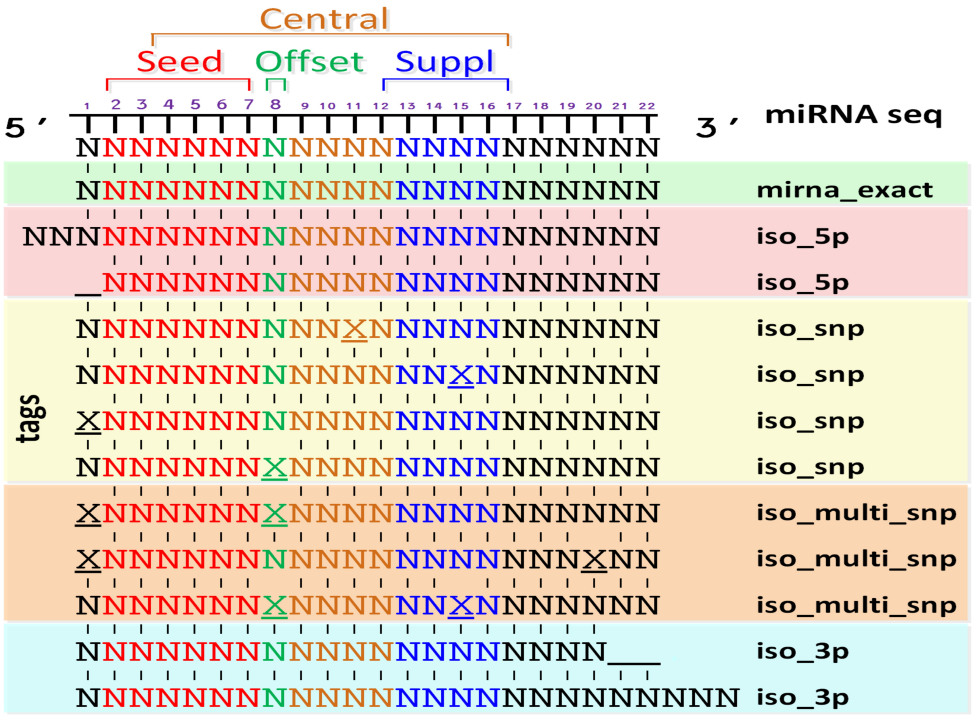
3rd Lab: miRNAs and isomiRs Analysis

PACIELLO Giulia
9-10 June 2016

IsomiR-SEA (1)

It is a novel tool that provides users with very accurate miRNAs expression levels and both isomiRs and miRNA-mRNA interaction sites precise classifications. Tags are mapped on the known miRNAs sequences thanks to a specialized alignment algorithm developed on top of biological evidence concerning miRNAs structure. Specifically, isomiR-SEA checks for miRNA seed presence in the input tags and evaluates, during all the alignment phases, the positions of the encountered mismatches, thus allowing to distinguish among the different isomiRs and conserved miRNA-mRNA interaction sites.



IsomiR-SEA (3)

Input Files:

- Tag file containing all the sequences that have to be analysed in order to detect miRNAs/isomiRs/conserved interaction sites
- miRNAs database containing all the known miRNAs sequences (in fasta format)

Output Files:

- miRNA Files: For each miRNA contained into the miRNAs database are reported the number of tags mapped on it and information concerning isomiRs and interaction sites.
- Tags Files: For each tag in the input file are reported different information as the number of identical tags mapped onto a given miRNA, the occurrence of isomiRs or conserved interaction sites.
    - unique_aligned_tags.txt: Stores all the tags that have been mapped on a unique miRNA
    - multi_aligned_tags.txt: Stores all the tags that have been mapped on more than a miRNA

IsomiR-SEA: unique_aligned_tags.txt

In Figure are reported 9 out of 23 fields contained in unique_aligned_tags.txt file

tag_index	tag_sequence	tag_quality	#count_tags	mirna_id	mirna_name	mirna_seq	seed_index	mirna_exact	iso_5p
1	UGGAGUGUGACAAUGGUGUUUG	è	2589502	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	1	0
3	UGGAGUGUGACAAUGGUGUUUGA	è	409656	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
4	UGGAGUGUGACAAUGGUGUUU	è	342764	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
6	UGGAGUGUGACAAUGGUGUUUGU	è	161746	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
7	UGGAGUGUGACAAUGGUGUU	è	137636	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
10	UGGAGUGUGACACUGGUGUUUG	è	80022	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
13	UGGAGUGUGCCAAUGGUGUUUG	è	58697	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
14	UACAGUACUGUGAUAACUGAA	è	53834	3222	>hsa-miR-10	UACAGUACUGUGAUAACUGAA	29	1	0
15	UGGAGUGUGACAAUGGUGU	è	51899	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
18	UAGCUUAUCAGACUGAUGUUG	è	38852	6221	>hsa-miR-21	UAGCUUAUCAGACUGAUGUUGA	60	0	0
19	UGGAGUGUGACCAUGGUGUUUG	è	37482	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
20	GUACAGUACUGUGAUAACUGAA	è	36180	3222	>hsa-miR-10	UACAGUACUGUGAUAACUGAA	29	0	1
28	CAACGGAAUCCCAAAAGCAGCU	è	26107	1243	>hsa-miR-19	CAACGGAAUCCCAAAAGCAGCUG	10	0	0
29	UAGCUUAUCAGACUGAUGUU	è	25595	6221	>hsa-miR-21	UAGCUUAUCAGACUGAUGUUGA	60	0	0
33	UAGCUUAUCAGACUGAUGUUGA	è	21508	6221	>hsa-miR-21	UAGCUUAUCAGACUGAUGUUGA	60	1	0
35	GGAGUGUGACAAUGGUGUUUG	è	20885	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	1
38	UCAGUGCACUACAGAACUUUGU	è	19271	11504	>hsa-miR-14	UCAGUGCACUACAGAACUUUGU	115	1	0
40	UCGGAUCCGUCUGAGCUUGGCU	è	19075	14235	>hsa-miR-12	UCGGAUCCGUCUGAGCUUGGCU	158	1	0
42	UGGAGUGUGACAAUGGUGUUUA	è	18905	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
43	UACAGUACUGUGAUAACUGA	è	18430	3222	>hsa-miR-10	UACAGUACUGUGAUAACUGAA	29	0	0
44	UGGAGUGUGACAAUGGUGUUUGAA	è	17270	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	246	0	0
46	GUACAGUACUGUGAUAACUGA	è	16819	3222	>hsa-miR-10	UACAGUACUGUGAUAACUGAA	29	0	1
47	ACCACAGGGUAGAACCACGGA	è	16592	3540	>hsa-miR-14	UACCACAGGGUAGAACCACGG	31	0	1

Seed Sequence Position:
- mirna_seq: 1-6
- Tag_sequence: variable

Each miRNA is identified by a unique mirna_id

If 1, the tag is aligned on a miRNA

If 1, the tag is aligned on a iso_5p

Lab Objectives (1)

Given *unique_aligned_tags.txt* file, write an ad-hoc python script that:

- For the different miRNAs identified by a specific *miRNA_id* calculates the total number of tags that have been perfectly aligned on it (Output file: mirna_id mirna_name #mapped_tags)
- For the different miRNAs identified by a specific *miRNA_id* calculates the the ratio between the number of tags mapped onto iso_5p and the number of tags mapped onto the exact miRNA sequence (Output file: mirna_id mirna_name ratio(iso5p/miRNA))
- For each tag that has been mapped on a iso_5p evaluates the shift of the seed sequence (Output file : tag_index tag_sequence mirna_seq shift:+/-x)
- Discuss the ratios iso_5p/miRNA calculated: Is there a prevalence of iso_5p or miRNA? How this finding impacts on miRNA targeting?

Example:

tag_index	tag_sequence	tag_quality	#count_tags	mirna_id	mirna_name	mirna_seq	seed_index	mirna_exact	iso_5p
20	GUACAGUACUGUGAUAACUGAA	è	36180	3222	>hsa-miR-10	UACAGUACUGUGAUAACUGAA	291	0	1
35	GGAGUGUGACAAUGGUGUUUG	è	20885	23475	>hsa-miR-12	UGGAGUGUGACAAUGGUGUUUG	2465	0	1
46	GUACAGUACUGUGAUAACUGA	è	16819	3222	>hsa-miR-10	UACAGUACUGUGAUAACUGAA	291	0	1
47	ACCACAGGGUAGAACCACGGA	è	16592	3540	>hsa-miR-14	UACCACAGGGUAGAACCACGG	315	0	1

Tag 20 → shift: +1 The seed sequence begins at index 2 on the tag instead at 1 as it is on miRNA
Tag 35 → shift: -1 The seed sequence begins at index 0 on the tag instead at 1 as it is on miRNA
Tag 46 → shift: +1 The seed sequence begins at index 2 on the tag instead at 1 as it is on miRNA
Tag 47 → shift: -1 The seed sequence begins at index 0 on the tag instead at 1 as it is on miRNA

Lab Objectives (2)

Given a tag file named *tags.txt* and a file storing all the known miRNAs sequences named *mirna.fa*, write a python program that:

- For each human miRNA (labeled as *hsa* within mirna.fa file) identifies the seed sequence and check its presence in the provided tags (stored in tags.txt file)
- If the seed is found in the tag in an allowed position (start seed position on the tag between index 0 and 4 included), prints on a file the following information:

  miRNA_name miRNA_sequence tag_sequence start_seed_position(on the tag)