MIRNA WARS
THE ISOMIRS MENACE

Lorena Pantano, P.h. D

https://lpantano.github.io
150 human liver samples with small RNAseq data

miRNAs

Gene regulation by imperfect complementary between seed region in miRNA and 3’UTR in the targeted RNA molecule.

isomiRs
GGGATGAGGTAGTAGGTTGTATAGTTTTTAGG
TGAGGTAGTAGGTTGTATAGTT
ATGAGGTAGTAGGTTGTATAGTTTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
TGAGGTAGTAGGTTGTATAGTT
What is the best way to analyze this data?
mirTOP
miRNA transcriptome open project
http://mirtop.github.io

repositories

Find a repository...
Type: All
Language: All

Repos

incubator
Where all ideas and discussions happen to lead to new repositories
R 3 4 Updated a minute ago

miRTOP.github.io
project for small RNA standard annotations
HTML 2 1 MIT 1 issue needs help Updated 4 days ago

mirtop
command lines tool to annotate miRNAs with a standard mirna/ismir naming
formatter mirna gff isomirs smallrna-seq
Python 6 13 MIT 17 issues need help Updated 5 days ago

simulator
first ideas and brainstorming for small RNA simulator
C++ Updated 25 days ago
Main projects:
- a format
- a way to compare
Format derived from GFF3: mirGFF3
a

## mirGFF3. VERSION 1.1
## source-ontology: miRBasev21 doi:10.25504/fairsharing.hmgte8
## COLDATA: sample1

b

Read=GATGAGG TAGTAGGTGTTG TAGTTGATAGTT -> UID=iso-24-5URPV39QFE
Read=ATGAGGTAGTAGGTGTTG TAGTTGATAGTT -> UID=iso-23-I0S31NSL0E
c

GGGATGAGG TAGTAGGTGTTG TAGTTGATAGTT TTAGG  Precursor
   TGAGG TAGTAGGTGTTG TAGTTGATAGTT
   ATGAGG TAGTAGGTGTTG TAGTTGATAGTT
   TGAGG TAGTAGGTGTTG TAGTTGATAGTT AA
   TGAGG TAGTAGGTGTTG TAGTTGATAGTT AA

   iso_5p:-1, iso_3p:+1
   iso_5p:+1, iso_3p:-1
   iso_add:2
   iso_5p:+1, iso_3p:-1, iso_add:2
d

\[ \text{iso_snv}_{\text{seed}} \quad \text{iso_snv}_{\text{central}} \quad \text{iso_snv}_{\text{central supp}} \]

T GAGGTA G TAGGTGTTGTA TAGTT iso_snv
iso_snv_{central_offset}
Python API: mirtop
Input:
seqbuster
miRge2.0
isomiR-SEA
sRNAbench
Prost!
mirGFF3
BAM

mirtop API
Importer
Converter
Manager
Compiler
Exporter

Output:
mirGFF3
isomiRs
VCF
Count matrix
What is the best way to analyze this data?
Samples suitable for benchmarking

Giraldez et al.

Tewari custom

Tewari synthetic
Different protocols

Library Preparation Protocols

Adapter Sequence

Library Method
(Simple)
TruSeq
CleanTag
NEBNext

(Detail)
TruSeq
CleanTag
NEBNext

Invariant

4N Random-End

4N

4N A
4N B
4N C
4N D
4N Xu
NEXTflex
Protocols

https://rnaseq.uoregon.edu/
<table>
<thead>
<tr>
<th></th>
<th>3 adapter</th>
<th>5 adapter</th>
<th>PEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>TrueSeq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEBNext</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>NextFlex</td>
<td>4N</td>
<td>4N</td>
<td>Yes</td>
</tr>
<tr>
<td>CleanTag</td>
<td>Methylphosphonate</td>
<td>2’0me</td>
<td></td>
</tr>
<tr>
<td>CATS</td>
<td>poly-T</td>
<td>TSO with rX</td>
<td></td>
</tr>
<tr>
<td>SMARTer</td>
<td>poly-T</td>
<td>TSO</td>
<td></td>
</tr>
</tbody>
</table>
Methods and Metrics
Sample

Library size = 20
Sample

miRNA 1 = 10

miRNA 2 = 10

Library size = 20
miRNA 1 = x10
isomiR 1.1 = x7
(match perfectly spike-in)

isomiR 1.2 = x3
(NOT match perfectly spike-in)

miRNA 2 = x10
isomiR 2.1 = x3
(match perfectly spike-in)

isomiR 2.2 = x5
(NOT match perfectly spike-in)

isomiR 2.2 = x2
(NOT match perfectly spike-in)

Library size = 20
There are 20 reads in total (lines), but 5 unique sequences (colors)
Sample

miRNA 1 = 10x
2 unique isomiRs

miRNA 2 = 10x
3 unique isomiRs

Library size = 20

isomiR 1.1 is 7/20=35% of the reads and 1/5=20% of the sequences
Library size = 20

miRNA 1 = 10
2 isomiRs

miRNA 2 = 10
3 isomiRs

isomiR 1.1 is 1/2 = 50% of the miRNA1 sequences

isomiR 1.1 is 7/10 = 70% of the miRNA1 reads
Sample

Library size = 20

miRNA 1 = 10
2 isomiRs

miRNA 2 = 10
3 isomiRs

PCT = isomiR 1.1 is 1/5=20% of the sequences

IMPORTANCE = isomiR 1.1 is 7/10=70% of the miRNA1 reads
Pipeline razers3 + mirtop
Data analysis - Filters
All sequences in a miRNA

Human?

Is the spike-in detected?

Any sequence in group mapped to other miRNA?

NO

NO

NO
Tewari-synthetic: library size
40 million reads and 85K different sequences.
Tewari-synthetic: spike-ins are the top sequence?
70% (692 being total detected miRNAs) of miRNAs with top sequence to be the expected one.
Most abundant type of isomiR for each miRNA without the spike-in as the most abundant

![Bar chart showing the abundance of different types of isomiRs for each miRNA sample.](chart.png)
Tewari-synthetic: Importance of the ‘isomiRs’ detected
isomiRs importance by type

**add3p**

**add3p + other**

**reference**

**shift3p**

**shift5p**

**shift5p shift3p**

**snp**

**snp + other**

**sample**

**IMPORTANCE**

- <0.1
- 0.1–1
- 1–5
- 5–10
- 10–20
- >20
- >50
From 85K to 4K unique sequences. 90% of sequences are removed.
isomiRs importance by type with pct > 1

importance

sample

IMPORTANCE

1–5

5–10

10–20

20–50

>50

add3p

add3p + other

reference

shift3p

shift5p

shift5p shift3p

snp

snp + other
Tewari-synthetic: Other tools
Tewari-synthetic: Other data
Tewari: synthetic vs human plasma
Tewari-synthetic: Shift type
% of isomiRs with shifting events

protocol

shift type

-10

-20

-30

-40

0-1

0-2

0-3

0-4

01

02

03

04

10

20

30

40

pct_cat

<0.1

1-5

10-20

0.1-1

5-10

20-50

clean

neb

tru

x4n

clean

neb

tru

x4n

clean

neb

tru

x4n

clean

neb

tru

x4n

clean

neb

tru

x4n

clean

neb

tru

x4n

clean

neb

tru

x4n
% of miRNAs with shift events

sample

protocol  clean  neb  tru  x4n

clean_tag_lab5  neb_next_lab1  neb_next_lab3  neb_next_lab4  neb_next_lab5  neb_next_lab9  tru_seq_lab1  tru_seq_lab2  tru_seq_lab3  tru_seq_lab5  tru_seq_lab6  tru_seq_lab8  tru_seq_lab9  x4n_a_lab5  x4n_b_lab2  x4n_b_lab4  x4n_b_lab6  x4n_c_lab6  x4n_d_lab1  x4n_nex_tflex_lab8  x4n_xu_lab5
Each miRNA generates a diversity of isomiRs:

- 90% contribute to < 1% of the miRNA abundance
- 10% enrichment on truncations at both sides
- Independent of pipelines and data sets

- 90% of miRNAs affected
- Custom 4N protocols perform worst
https://github.com/miRTop/incubator/tree/master/projects/tewari_equimolar

https://github.com/miRTop/mirGFF3

https://github.com/miRTop/mirtop
Thomas Desvignes
Phillipe Loher
Karen Ellbeck
Bastian Fromm
Gianvito Urgese
Isidore Rigoutsos
Michael Hackenberg
Ioannis S. Vlachos
Marc K. Halushka

Jeffery Ma, Jason Sydes, Yin Lu, Ernesto Aparicio-Puerta,
Shruthi Bandyadka, Victor Barrera, Peter Batzel,
Rafa Allis, Roderic Espin

Kieran O’Neill, Eric Londin, Aristeidis G. Telonis,
Elisa Ficarra, John H. Postlethwait,
MiRNA WARS: THE NEW HOPE
Thank you!