New computational and single-cell transcriptomic approaches in microRNA biology

Marc Friedländer
3rd Annual Danish Bioinformatics Conference
Science for Life Laboratories/Stockholm University
Todays presentation

0. Introduction to miRNA biology

1. miRTrace reveals the organismal origin(s) of miRNA-seq data

2. sequencing of miRNAs in single cells
0. Introduction to miRNA biology
Non-coding RNA genes

- **nematode:** 20,000 protein coding genes
- **human:** 20,000 protein coding genes
Non-coding RNA genes

- **nematode:**
  - 20,000 protein coding genes
  - 6,000 non-coding RNA genes

- **human:**
  - 20,000 protein coding genes
  - >25,000 non-coding RNA genes
Non-coding RNA genes

- nematode:
  - 20,000 protein coding genes
  - 6,000 non-coding RNA genes

- human:
  - 20,000 protein coding genes
  - >25,000 non-coding RNA genes

- some non-coding RNAs have ‘house-keeping’ functions, like rRNAs and tRNAs
- other have regulatory functions, like long non-coding RNAs and miRNAs
Importance of miRNA regulation

- small RNAs that regulate expression of protein coding genes
- appear to be present in (virtually) all multicellular animals
- knock-outs are embryonically lethal
- regulate 30-60% of coding genes
- involved in numerous biological processes including:
  - cell identity
  - development
  - disease states
miRNA biogenesis and function

Biogenesis: two cuts

Function:

- inhibition of target mRNA translation
- reducing target mRNA stability
- localization of target mRNAs to P-bodies
miRNA sequencing

Library preparation:
- similar to ordinary RNA-seq
- contains gel purification step

Computational challenges:
- snap-shot of RNA turn-over
- short (~22 nts) sequences
- many multi-mapping sequences
1. miRTrace reveals the organismal origins of miRNA sequencing data
Tracing organismal origin(s) of samples

*Numerous applications:*
- criminal forensics
- clinical parasitology
- food quality control
- research laboratory contaminations

*Traditional methods:*
- rDNA analyses
- marker genes
miRNA evolution

- Protein machinery present in last eukaryotic common ancestor
- Likely convergent evolution in animals and plants
- miRNA genes emerge continuously in evolution
- Once fixated, they are rarely secondarily lost
- They are highly conserved in sequence
miRTrace concept

- catalog clade-specific miRNAs
- analyze miRNA-seq data, noting sequence matches to catalog
- report composition on clade-specific miRNAs

Advantages over traditional methods:
- binary analysis (miRNA either present or completely absent)
- works for species not yet studied at the sequence level
- uses information from dozens of distinct loci
- method simple and requires no expert knowledge or models
miRTrace accurately identifies clade origins
miRTrace sensitivity

Single-cell miRNA sequencing
miRTrace sensitivity

Parasite RNA in blood
miRTrace sensitivity

In silico contaminations
Cross-clade contaminations in public data

Primate-specific contaminations in data from model species
Possible source of contaminations

Standard Illumina read for sRNA-Seq

Sample multiplexing

Sequencing

Demultiplexing

Primate-specific miRNA read
Rodent-specific miRNA read
Other read
Cross-clade contaminations in in-house data

Cross-species contamination from index mis-assignment
Cross-clade contaminations in in-house data

Cross-species contamination from index mis-assignment
Cross-clade contaminations in in-house data

Cross-species contamination from index mis-assignment
Cross-clade contaminations in in-house data

Cross-species contamination from index mis-assignment
Effects of cross-clade contaminations

In silico contaminations from 0-100%
Effects of cross-clade contaminations
Effects of cross-clade contaminations
Effects of cross-clade contaminations

Effect of contamination on novel miRNA prediction (human contamination)

- False negatives (out of 590)
- False positives
Cross-clade contaminations in public data

Primate-specific contaminations in data from model species
Cross-clade contaminations in public data

Primate-specific contaminations in data from model species
miRTrace summary and outlook

- miRTrace accurately reveals organismal origin(s) of miRNA-seq data
- contaminations prevalent in public data, but mostly at trace levels
- some contaminations can be removed even after sequencing
- better catalogs of miRNA phylogeny would strengthen method
- small RNAs from unicellular organisms could be included
- our method is implemented as an overall quality control tool
2. sequencing of miRNAs in single cells
Variability between cells

What we observe in our data:

Biological explanation:

OR
miRNA single-cell sequencing

- In total 48 mouse embryonic stem cell
- unique molecular identifiers (UMIs)
- we sample around 1% of the miRNA molecules (*poll* rather than *census*)

Figure from Faridani et al., Nature Biotech 2016.
miRNA variation of expression in single cells
miRNA variation of expression in single cells

miRNAs involved in pluripotency

miRNA of unknown function
miRNA ‘arm’ selection

For some miRNAs, both arms can enter Argonaute and guide repression
miRNA ‘arm’ selection in single cells

Arm bias:

significant

not significant
miRNA adenylation
miRNA adenylation

Some miRNAs are adenylated, which change their stability.
miRNA adenylation in single cells

**Adenylation bias:**
- significant
- not significant

![miR-92a-3p histogram]

- untemplated A/total reads
  - 1611/6376
miRNA single-cell sequencing: conclusions and outlook

Conclusions:
- overall, miRNA content in single cells reflect bulk data
- however, adenylation can differ radically between cells

Outlook:
- correlating with gene expression can suggest proteins responsible
- improving method to give higher miRNA yield
Acknowledgements

*Friedländer group*
Wenjing Kang
Sebastian Mackowiak
Marcel Sauerbier
Inna Biryukova
Franziska Bonath
Yrin Eldfjell

*Karolinska Institutet*
Omid Faridani
Rickard Sandberg

...and thank you for your attention!
Clade-specific miRNA families

59

16

21

7

9

26

61

4

8

175

62

6

33

67