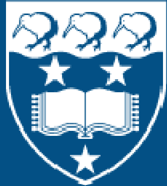


THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

MicroRNAs

Sequencing, analysis … and then what?

Nicole Cloonan - 4th July 2017

#UQWinterSchool #Bioinformatics #GroupTherapy

Woohoo! I’m going to do a miRNAseq project!

Here are some things to think about.

OMFG – my career is over.

The five stages of miRNA analysis

Denial

OMG – I have started an amazing project, this will get me a Nobel Prize for sure!

Ancient Greek word

"little"

Ancient Greek word

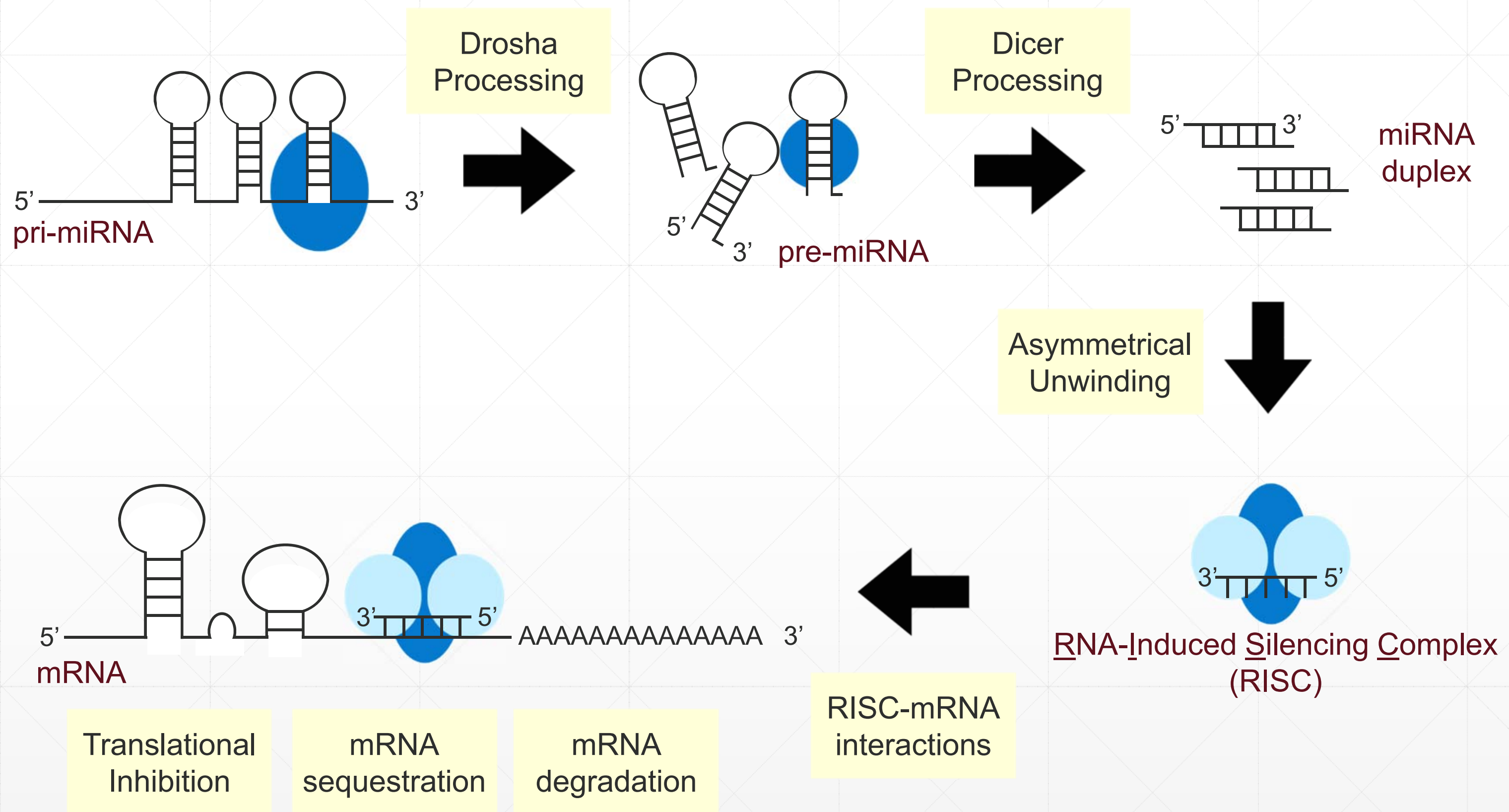
"bastard"

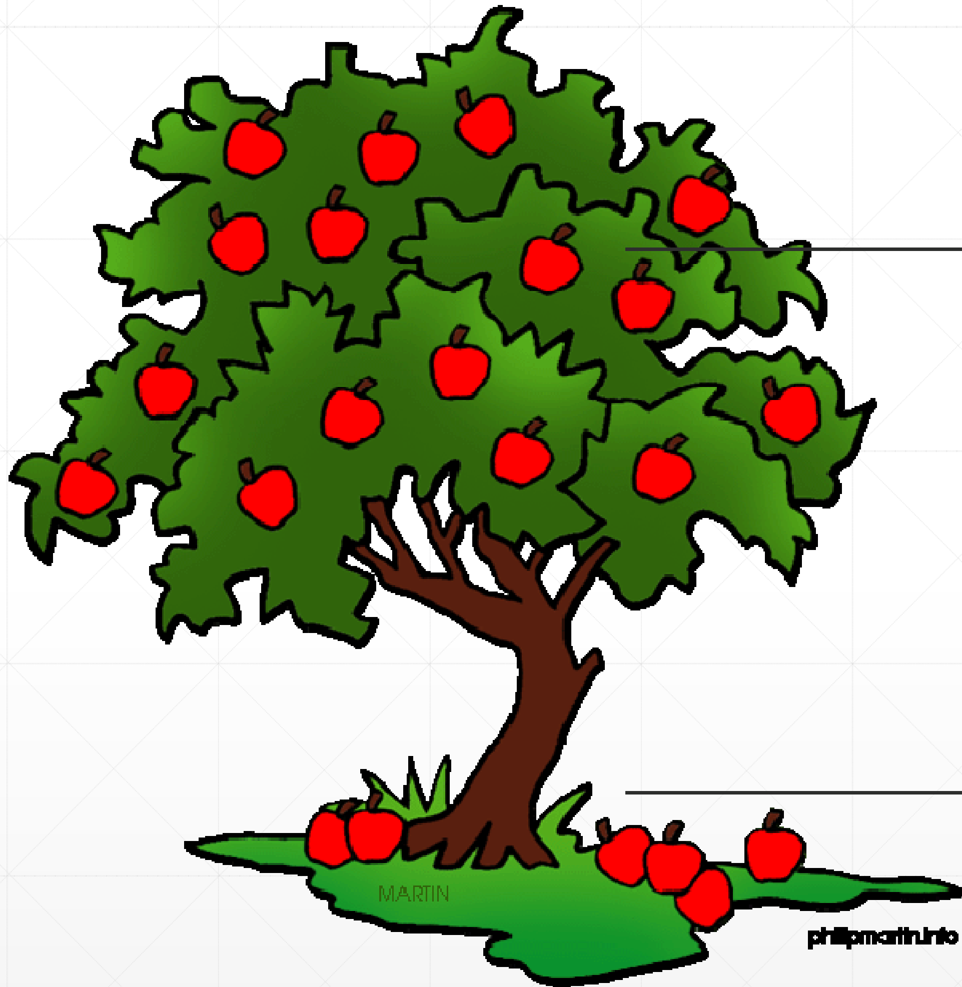
microRNA

noun mi·cro·RNA \ˌmī-krō-ˌär-(ˌ)en-ˈā\

variants: **miRNA** *or less commonly* **micro–RNA**

miRNA biogenesis





Low Hanging Fruit

Cheap and easy to sequence (~$80 per sample)
A natural fit for short read sequencing technologies
Fantastic possibilities as biomarkers (stable)
Interesting possibilities as therapeutics
Established modes of action
Enables lazy department store science

Ground Fruit

Flawed, unpalatable, and a probably a little smelly

The five stages of miRNA analysis

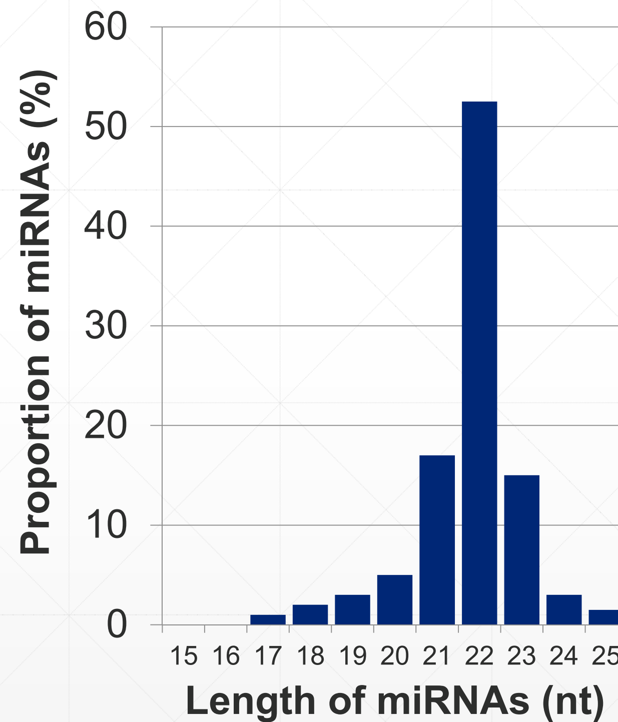
Anger

What do you mean we can't align these things well?

Anger – the project is not as easy as your supervisor made you think it was

```
miR-17-5p : CAAAGUGCUUACAGUGCAGGUAGU
miR-20    : UAAAGUGCUUAUAGUGCAGGUAG-
miR-106a  : AAAAGUGCUUACAGUGCAGGUAGC
miR-106b  : UAAAGUGCUGACAGUGCAGAU---
miR-93    : -AAAGUGCUGUUCGUGCAGGUAG-
miR-18    : UAAGGUGCAUCUAGUGCAGAUA--

miR-19a   : UGUGCAAAUCUAUGCAAAACUGA-
miR-19b-1 : UGUGCAAAUCCAUGCAAAACUGA-
miR-19b-2 : UGUGCAAAUCCAUGCAAAACUGA-
```

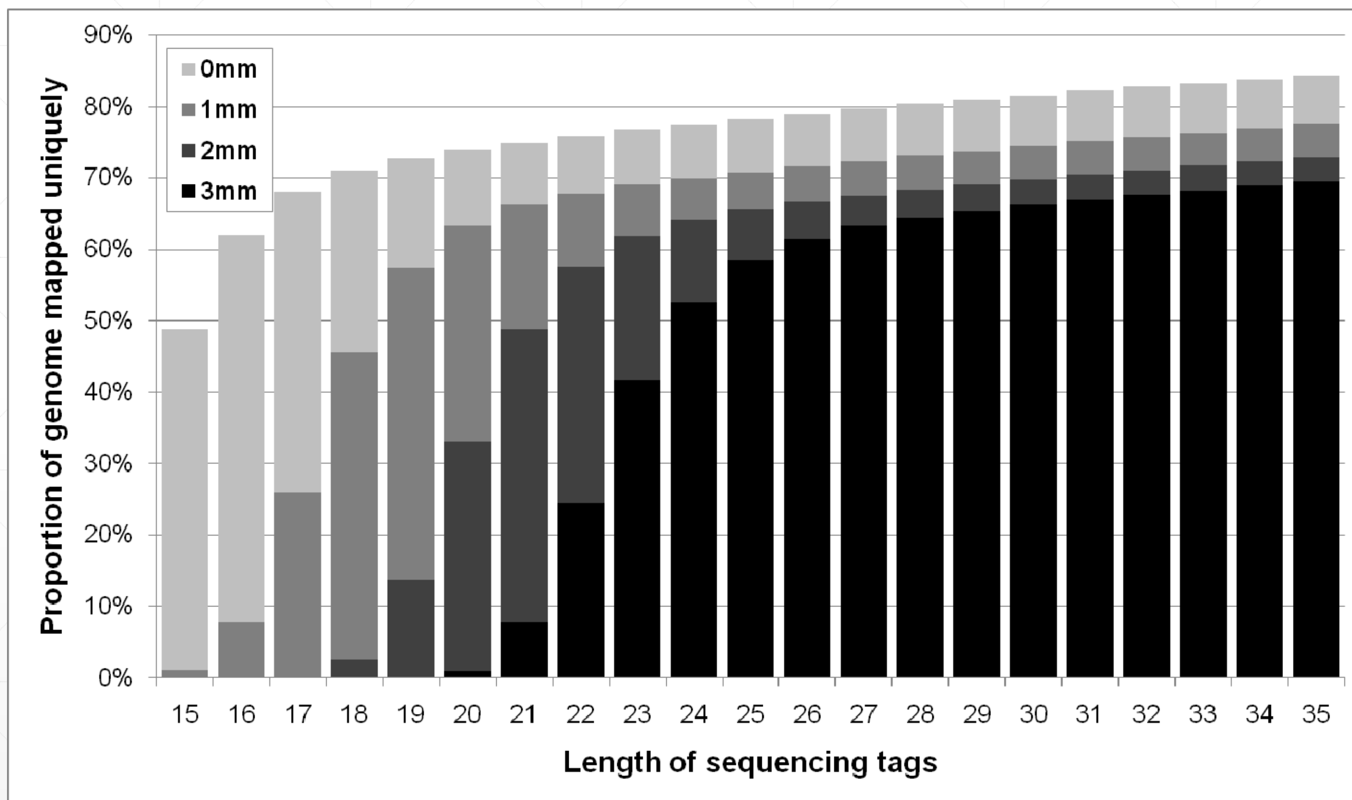


The two laws of repeats

1. It is impossible to resolve repeats of length L unless you have reads longer than L.
2. It is impossible to resolve repeats of length L unless you have reads longer than L.

Gleefully stolen from @torstenseemann

Aligning to the genome



Allowing errors in alignments

- Your reference genome will contain polymorphisms
- Your sample will contain amplification or sequencing artefacts
- Your sample may contain RNA editing
- Unique alignment is not the same as specific alignment
- Therefore, you must allow at least as many errors per alignment as you are expecting in your data

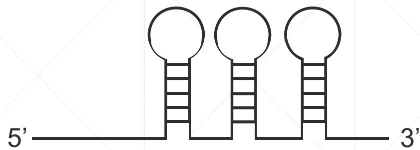
Watch out for your aligner!

- Make sure your aligner and/or alignment parameters allow proper mapping of small RNAs.

miRNA size = 22nt caaagugcuuacagugcaggua

Seed size = 16nt

Alternatives to genome alignment



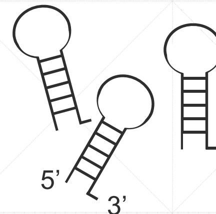
pri-miRNA transcript

Could also be pre-mRNA transcript with miRNAs in the introns

1000s nt long

Often not fully defined

Will miss novel miRNAs



pre-miRNA hairpins

80-200 nt long

Will miss novel miRNAs

Mature miRNAs

18-25 nt long

Will miss novel miRNAs

Will not detect isomiRs

Best curated resource for miRNAs is currently miRbase (**http://miRBase.org**)

The five stages of miRNA analysis

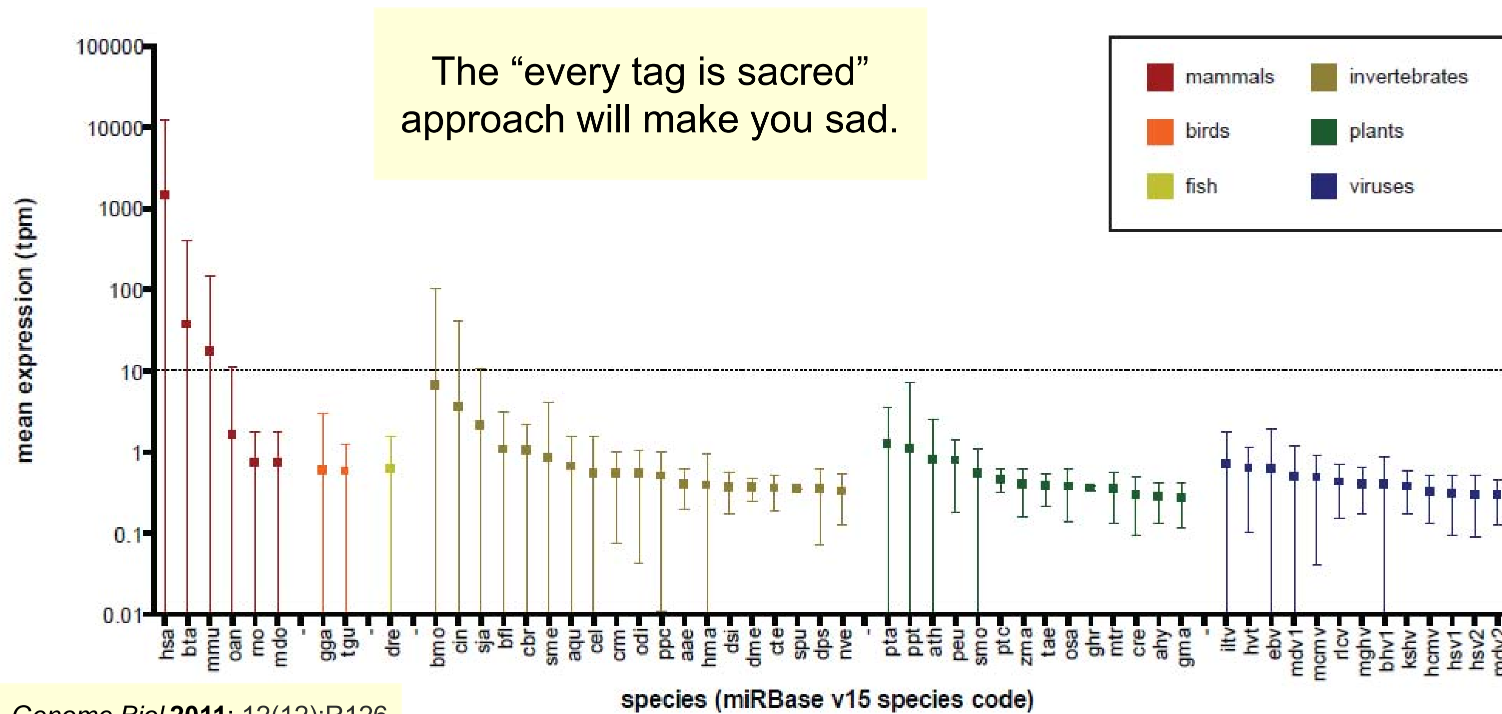
Bargaining

Can we please have some more replicates?

Bargaining – maybe I can still get useful data if I do the analysis well

1. What miRNAs are expressed in my sample?
2. What miRNAs are different between my samples?

What miRNAs are expressed?



Cloonan *et al. Genome Biol* **2011**; 12(12):R126

What miRNAs are different?

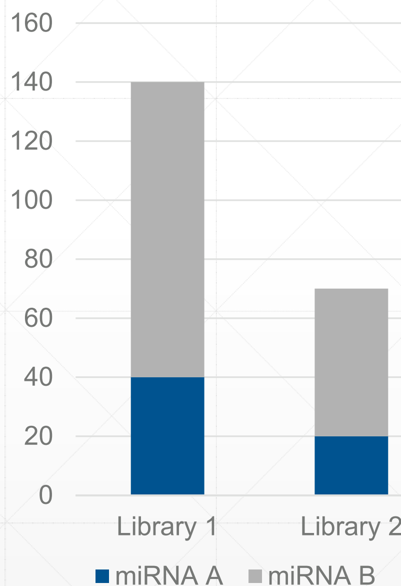
- You absolutely CANNOT answer this question without replicates
  - All biological systems are variable – it's an evolutionary NEED
  - Understanding biological variations means no wasted time chasing false positives
  - Your question is actually whether the variation between your groups is bigger than the variation within your groups
- You must have independent biological replicates
  - Do not fall into the "reducing variation" trap
  - Variation is your friend, not your enemy – you want to MAXIMIZE your biological variation
  - Using technical replicates makes your p-values WORTHLESS
- If you can't afford replicates – then you can't afford the experiment.

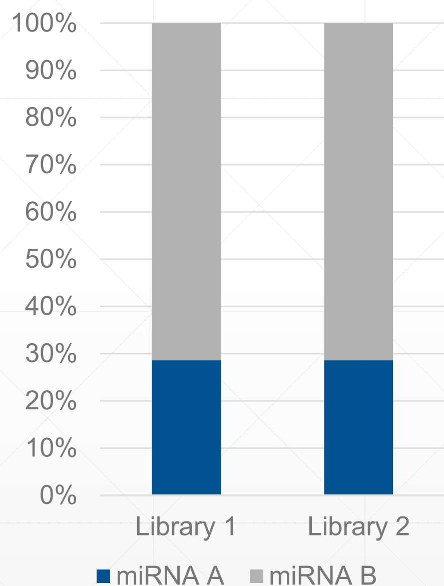
Comparing groups - Normalization

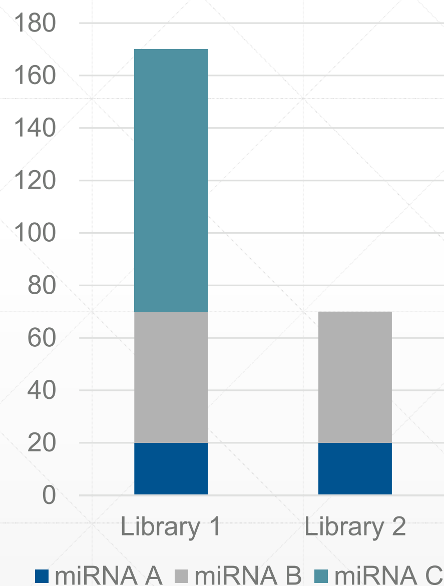
- "Housekeeping Genes"
  - Could we all please agree that this is not a thing and to stop bringing it up?
  - You shouldn't even be doing this for qRT-PCR let alone sequencing
- Spike-Ins
  - One potential method of normalization, but be careful
  - There is error when measuring any individual transcript, when you "correct" with a spike-in, you multiply the spike-in error with the error of each transcript
  - A spike in added as a percent of the RNA instead of a percent of the cells will mask any global changes in RNA production or composition.
- Data-driven normalization (eg. Quantile normalization)
  - Be careful – relies on the assumption that the vast majority of genes/miRNAs will not change
  - Only a few thousand data points for these approaches to model – check software assumptions

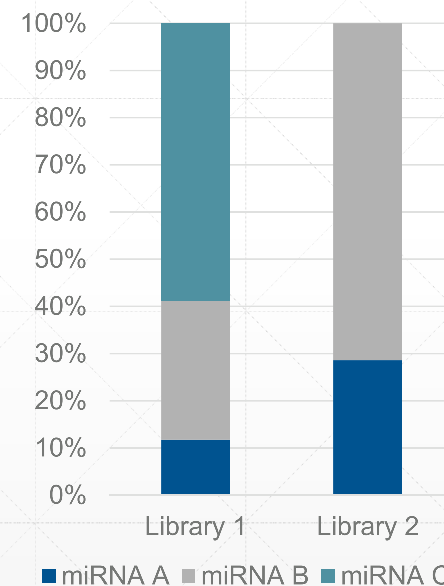
Scaling to Library Depth

TMM – Trimmed Mean of M Values





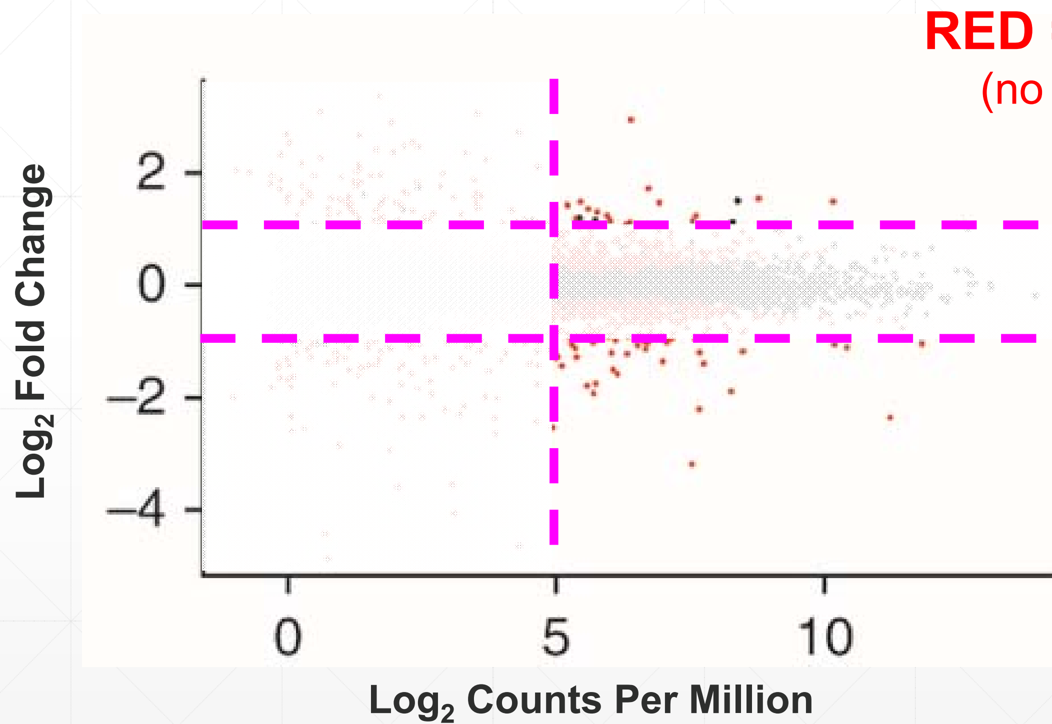




Robinson & Oshlack *Genome Biol* **2010**; 11(3):R25

What miRNAs are different?

RED = statistical significance
(no replicates = no p-values)



The five stages of miRNA analysis

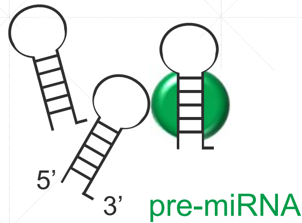
Depression

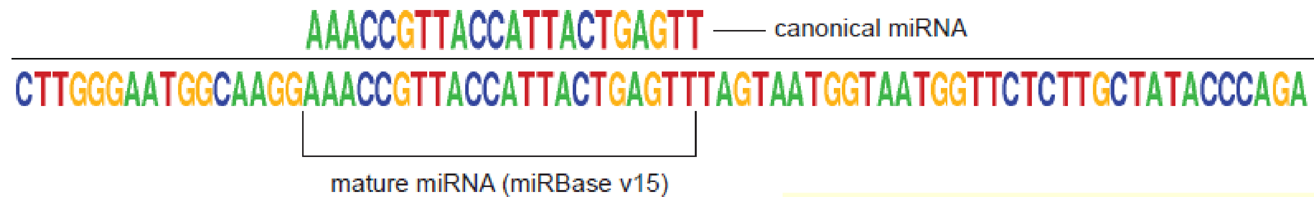
OMFG – do these miRNA problems never end?

Depression - you have a list of miRNAs but realise that no one cares

- If you are doing a biomarker study, then you need to demonstrate utility
- If you are doing a mechanistic study, then you need to demonstrate function

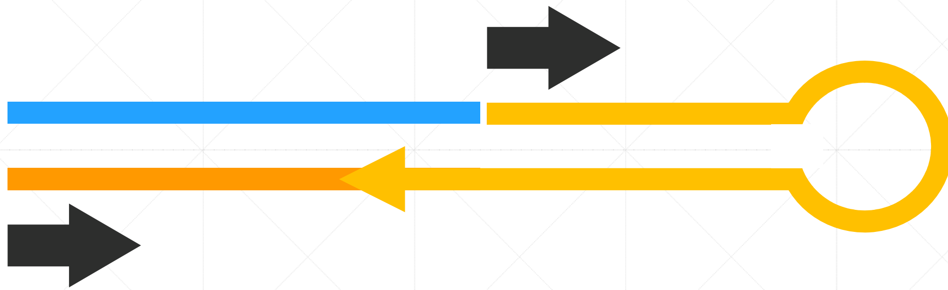
IsomiRs and biomarker validation

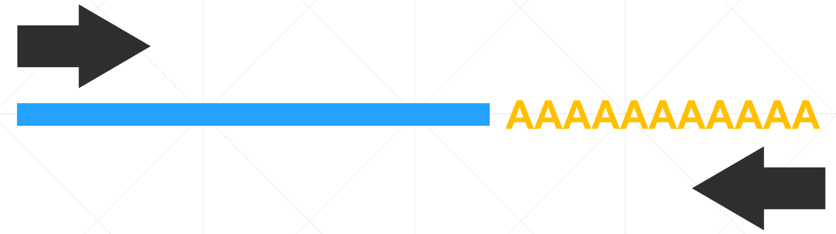




Cloonan *et al. Genome Biol* **2011**; 12(12):R126

qRT-PCR and IsomiRs





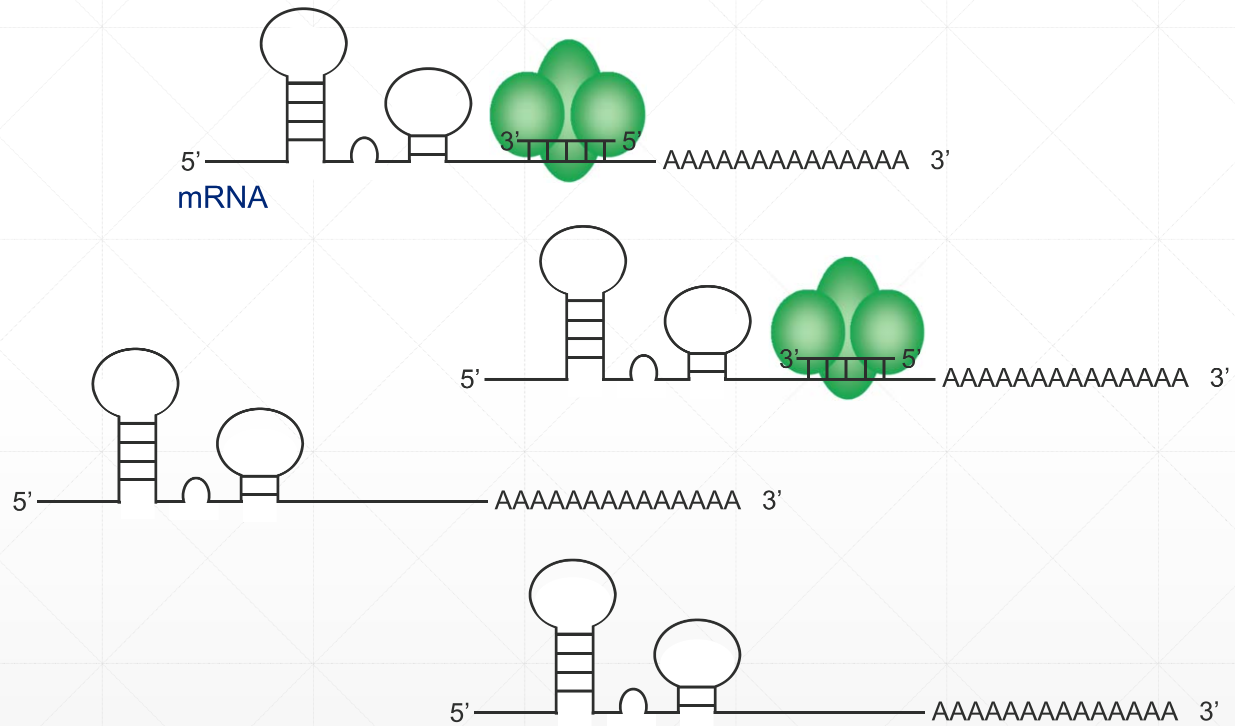
This method is isomiR specific

This method is isomiR agnostic

Mechanism - What is a target anyway?

- miRNAs can bind to mRNAs without inducing a repressive effect
  - This happens with transcription factors due to inherent non-functioning sites, or limiting co-factor concentrations
- Full or partial repression of a protein/mRNA does not necessarily alter the trajectory of a cell
  - Different molecules have different tolerances of variability in expression
- Full or partial repression does not mean a change in protein/mRNA levels
  - Long half lives, stoichiometry

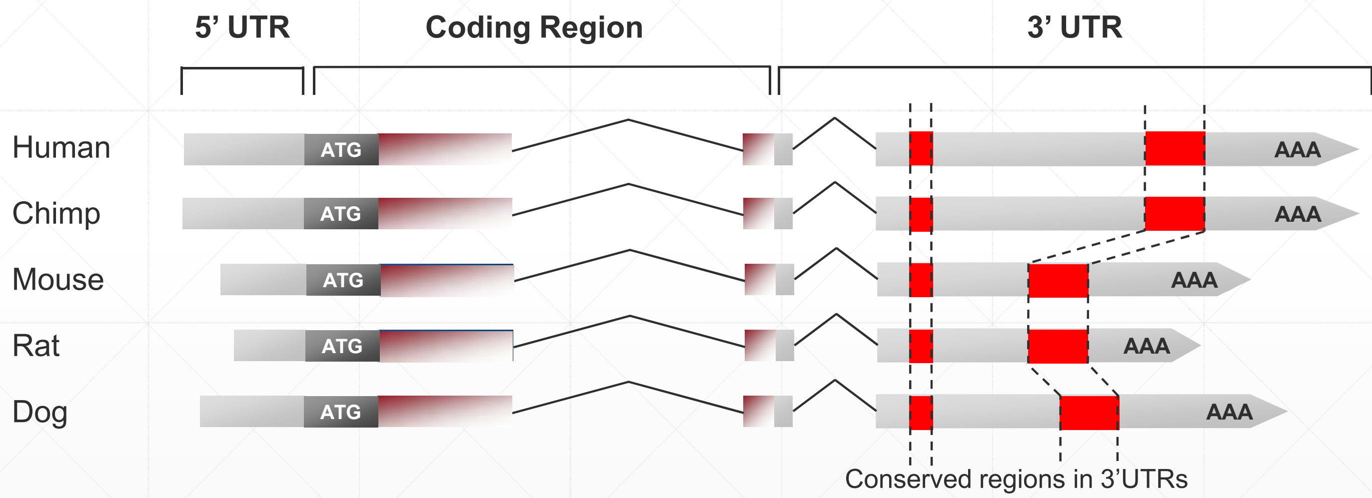
Stochiometry of miRNA Interactions



Cloonan *et al.*, *Genome Biology*, **2008** 9:R127

Bioinformatics methods for target PREDICTION

Assumptions have to be made to achieve any kind of specificity

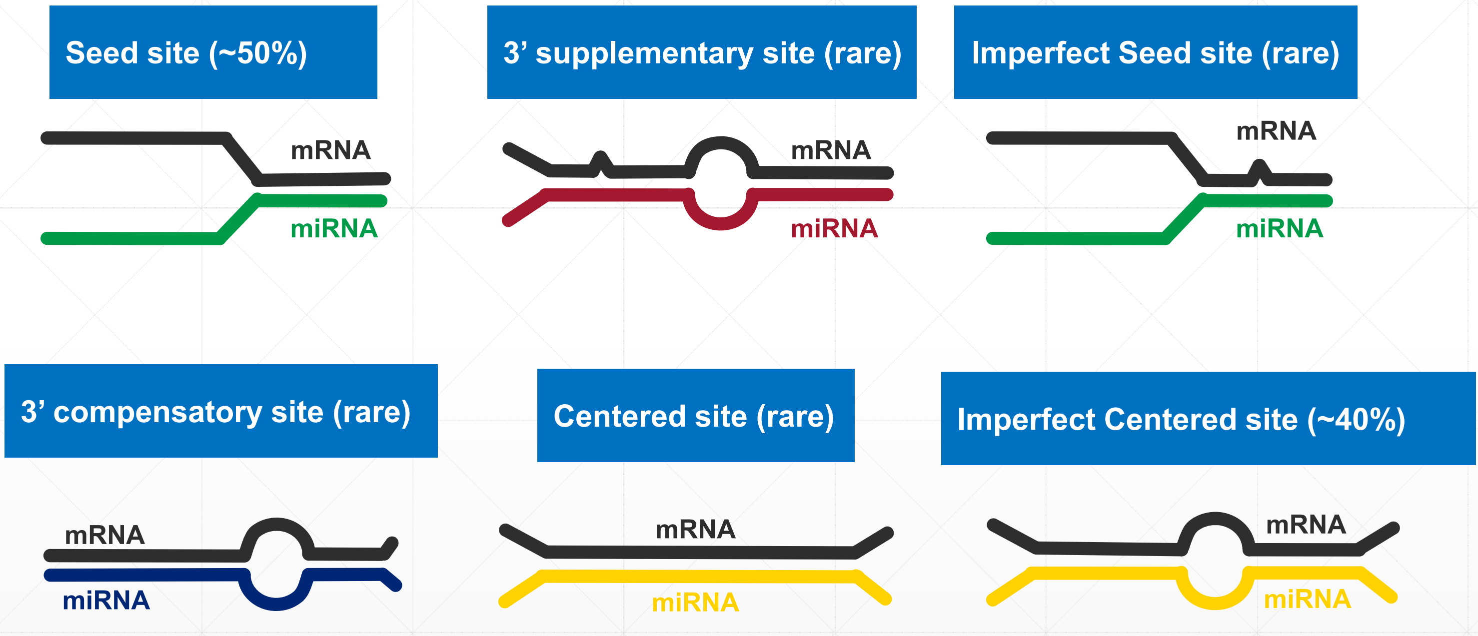


Best have 50% false positive rates, and at least 40% false negative rates

The dominance of the seed site

- Nucleotides 2-8 in the 5’ end of the miRNA are called the “seed” site.
  - Fun fact: first miRNA interactions were not seed sites
  - Fun fact: seed sites have the most predictive value. This is NOT the same thing as being the most functional
- Literature bias towards seed sites because experimental validation focuses around those most likely to give the “best” results for the lowest time/cost investment

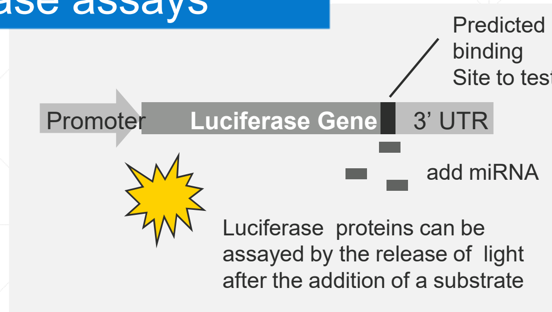
Modes of miRNA binding

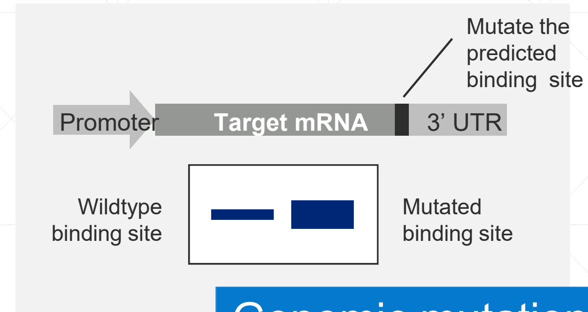


Adapted from Brennecke *et al.*, (2005) PLoS Biology **3**:e85

Target detection methods

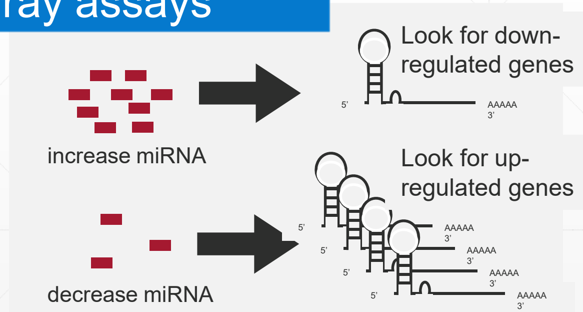
Luciferase assays

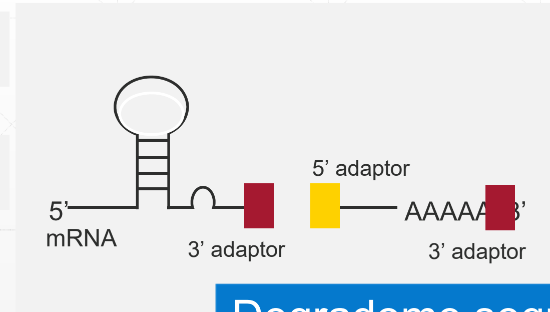




Genomic mutation

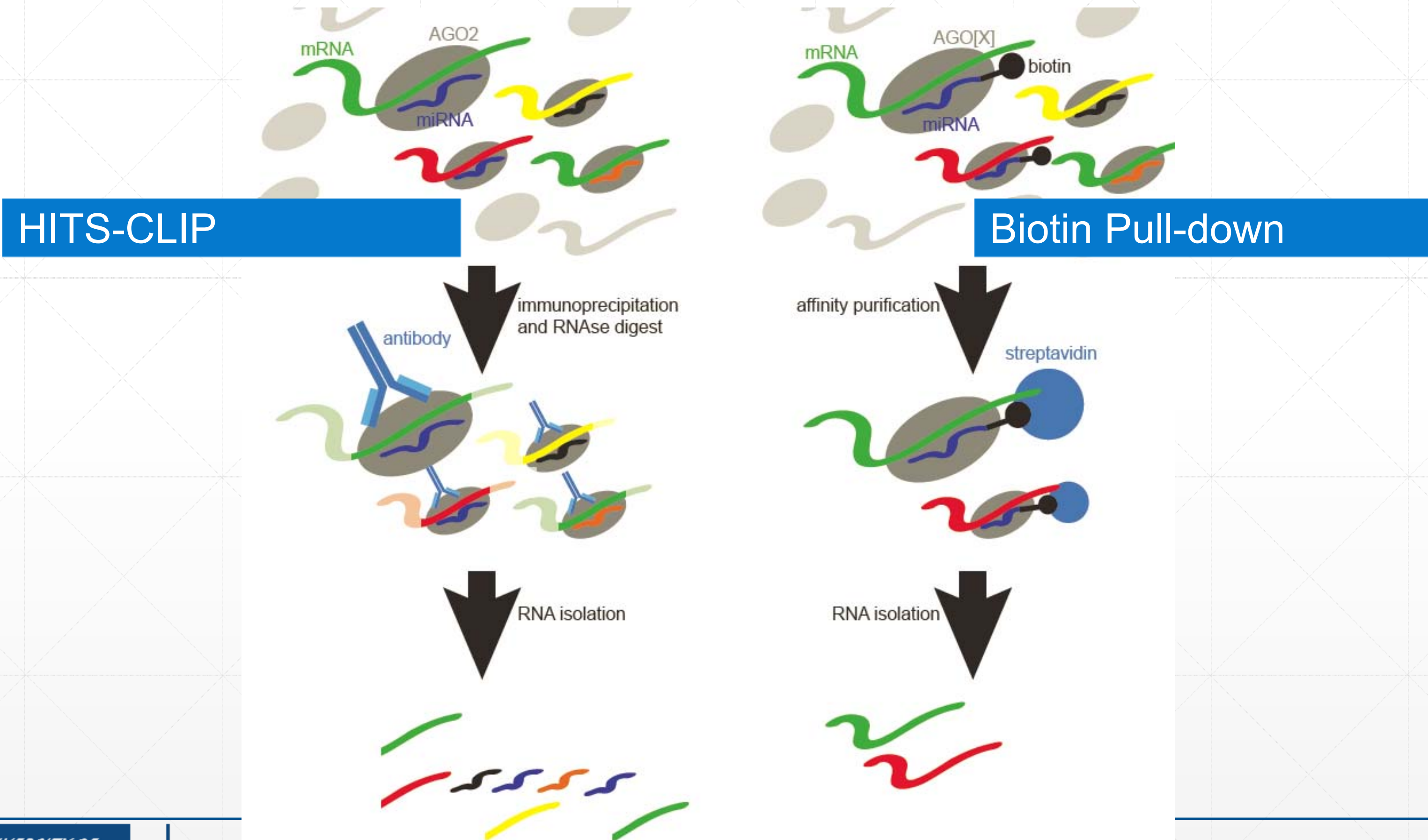
Microarray assays





Degradome sequencing

Target detection methods



Martin *et al.*, *Genome Biology*, **2014** 15:R51

The five stages of miRNA analysis

Acceptance

OK, it's not as amazing as I thought, but it's probably still possible to write a decent paper.

Conclusions

- Seriously – use replicates. No, really. Do it.
- Check your assumptions. No, really. Do it.
- Methods matter, understand what each method tells you... and what it DOESN'T

MicroRNAs can be a fun project, but are not an easy project. Exercise caution and don't over interpret your results.

Last Word…

99.5% of people with a biology PhD will not retire while still in a research role. Hard work and talent are not sufficient to guarantee a long career. Your time in research may be short, but your contribution to the literature is FOREVER.

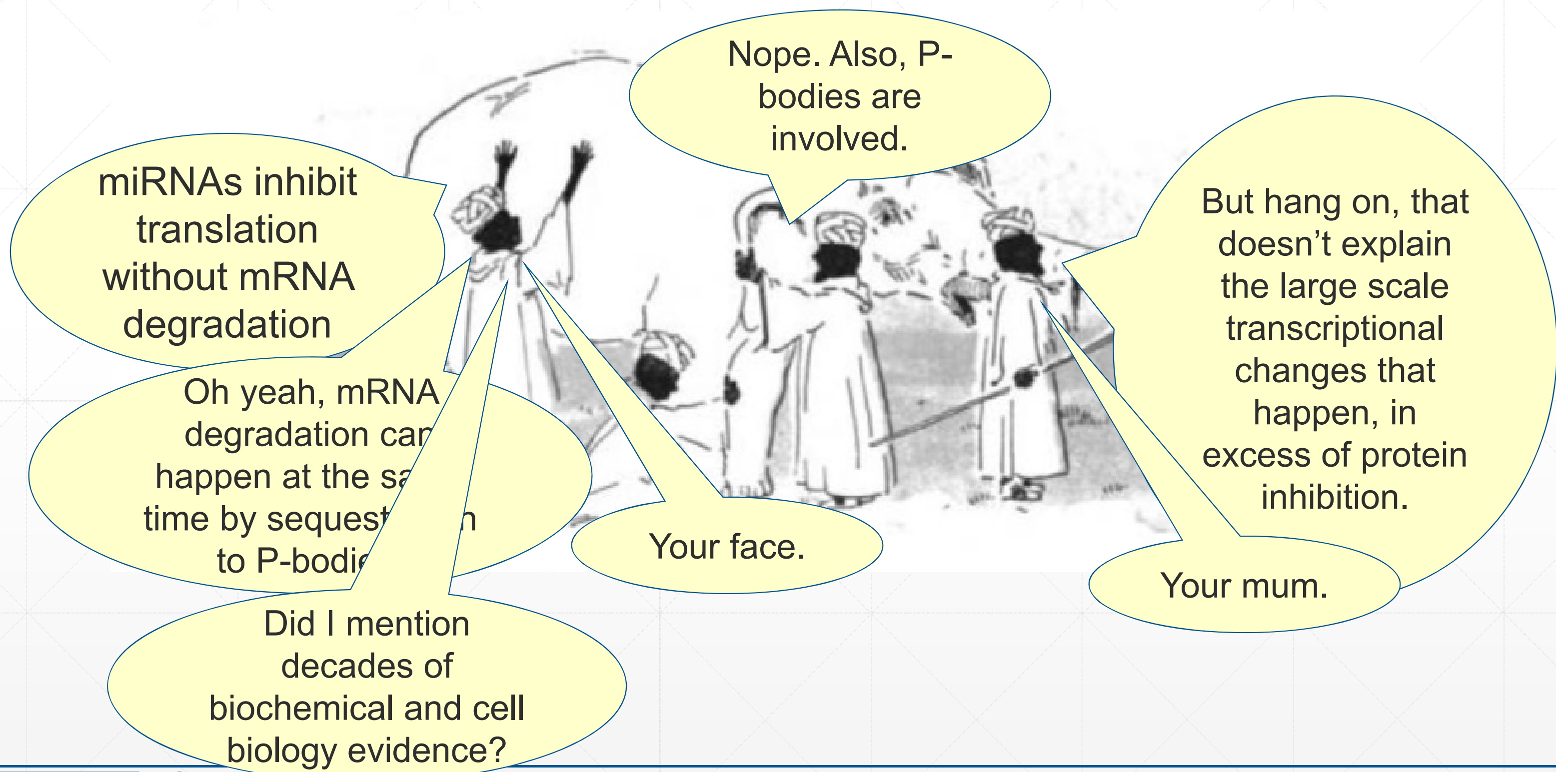
So take the time to do GOOD science.

Our Time is Up!

@ncloonan

n.cloonan@auckland.ac.nz

A parable of miRNA function



THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

SCIENCE
SCHOOL OF BIOLOGICAL SCIENCES