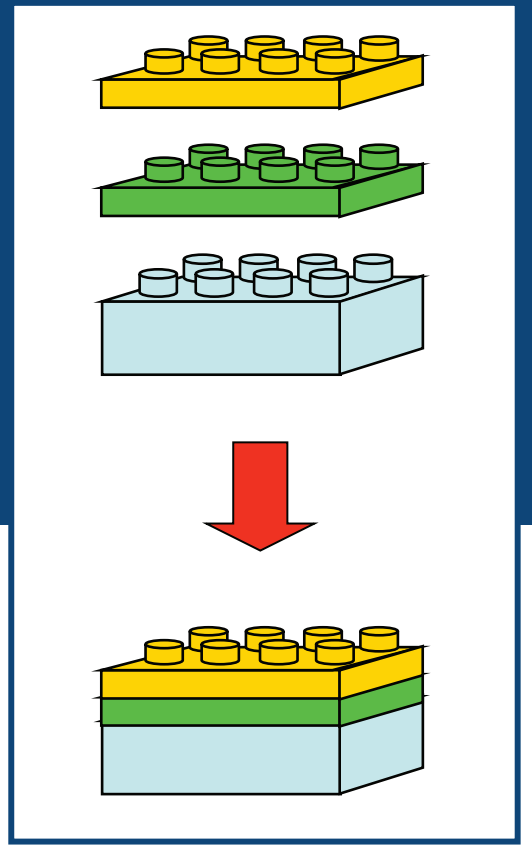
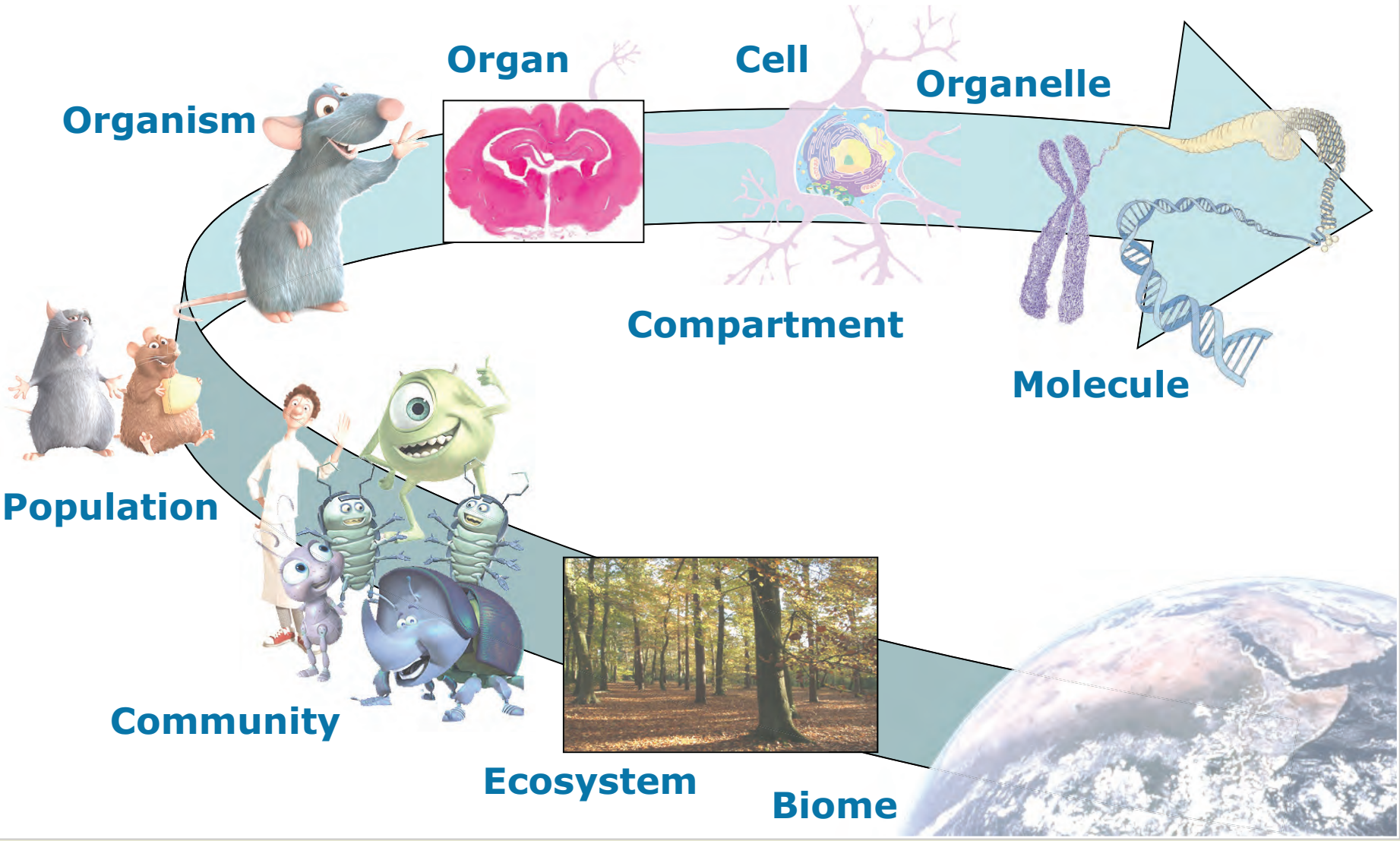
Modular organization and composability of RNA



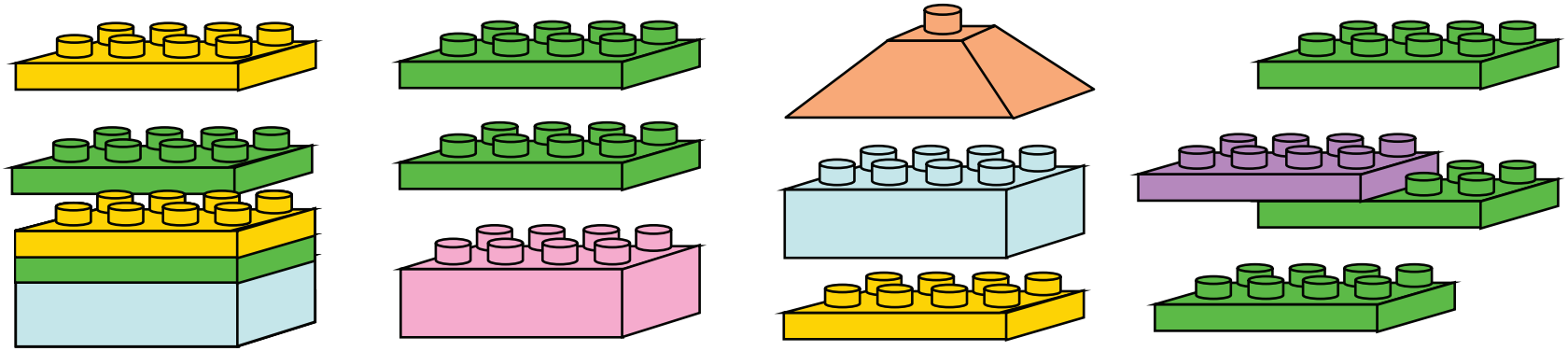
Miler T. Lee
Laboratory of Junhyong Kim / Univ. of Pennsylvania
DOE CSGF Annual Conference
16 July 2009

Penn
Genomics & Computational Biology

Life is organized



Organization achieved via composability, modularity



Composability

At some level of abstraction, the system consists of a series of smaller parts that serve as building blocks

Modularity

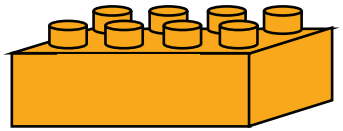
These building blocks are independent units that can be recombined to form functionally different systems

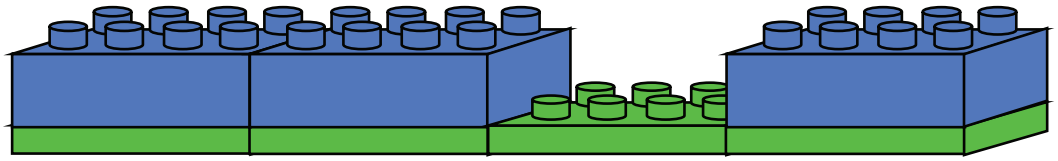
Example of modular architecture

Individual components separately inserted into the system

(Abstractly) framing modularity in RNA terms

Building an RNA molecule...





Composite RNA

Q1 Does the orange block change after insertion?
Do the blue blocks make a difference in this?

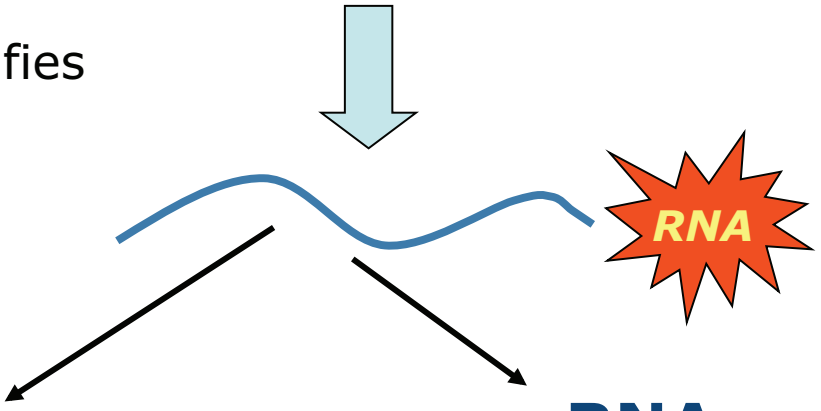
Q2 Are the blue blocks changed by the insertion?
Does it even matter if they are changed?

If the orange block is a modular component, the answer is **NO** to both.

What is RNA?

GENE (DNA)

Gene sequence specifies RNA sequence



mRNA
messenger RNA

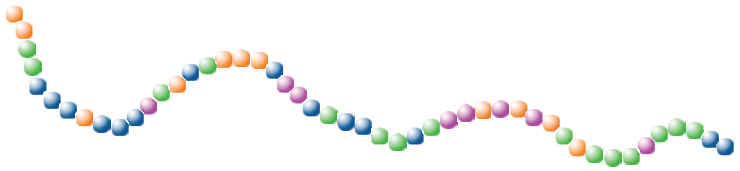
- Intermediate instruction carrier specifies a protein product
- Sequence fidelity is necessary to correctly specify the protein

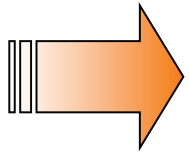
ncRNA
non-coding RNA

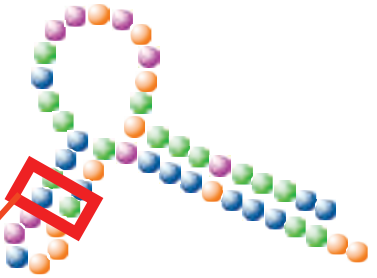
- End product with catalytic function (e.g., tRNA, ribosomal)
- Sequence fidelity is necessary to correctly specify the **structure** of the molecule

RNA structure

▶ Sequence determines structure

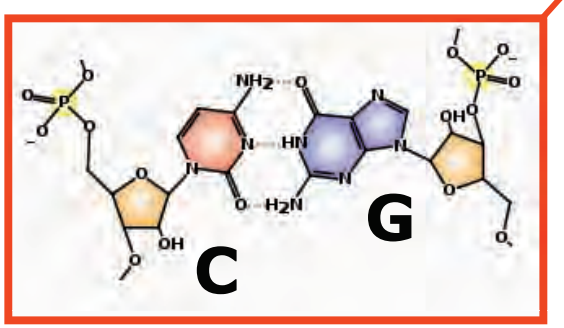






A=U
C≡G
G≈U

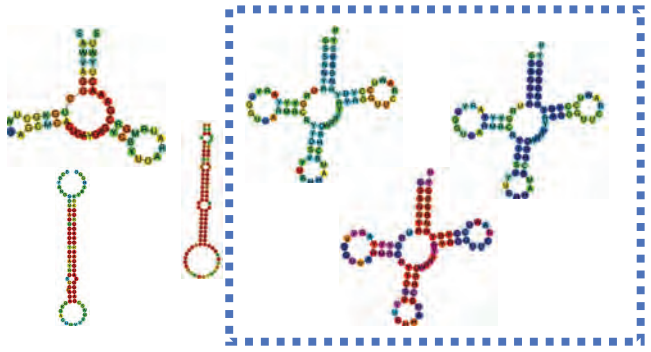
Linear sequence of nucleotides
{**A**, **C**, **G**, **U**}



Pairwise free-energy-minimizing interactions form a folded **"secondary" structure**

▶ Structure determines function

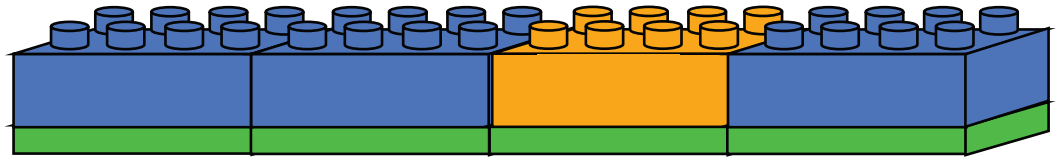
Different RNAs sharing common structure but not necessarily sequence tend to share common function

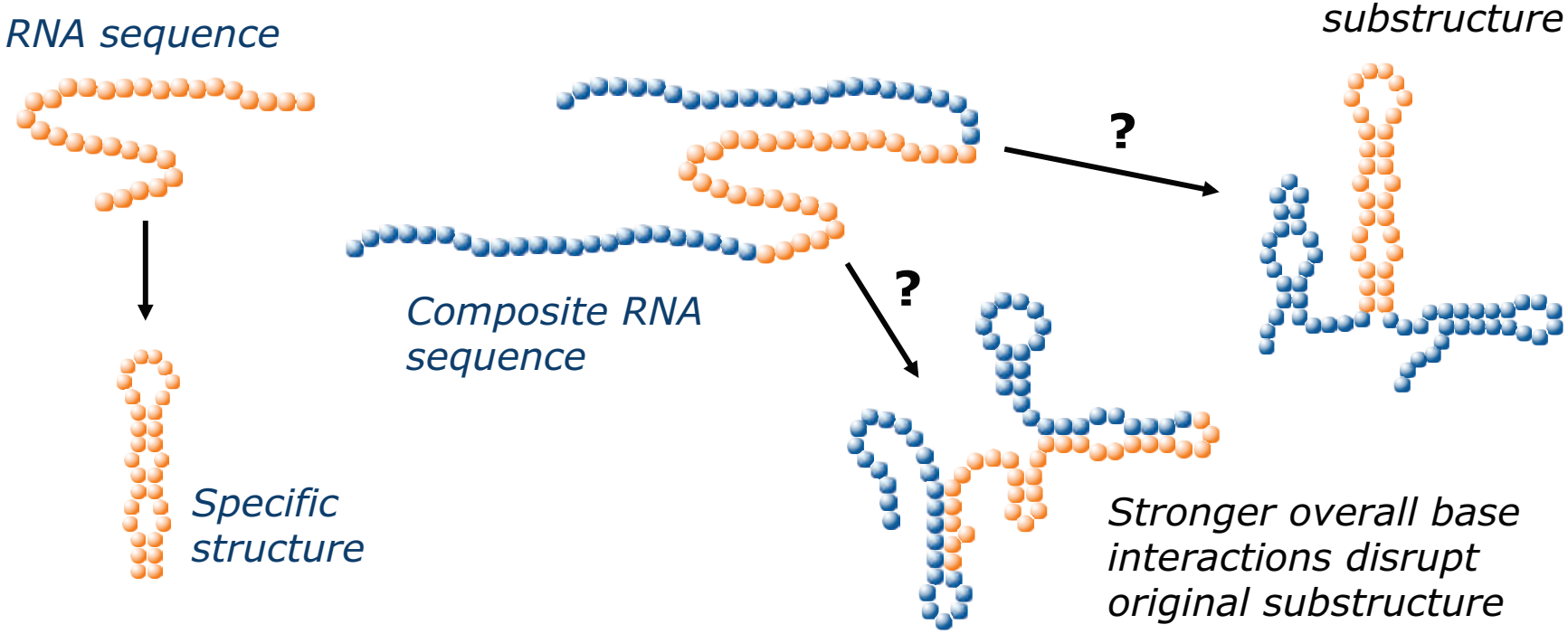


tRNA family

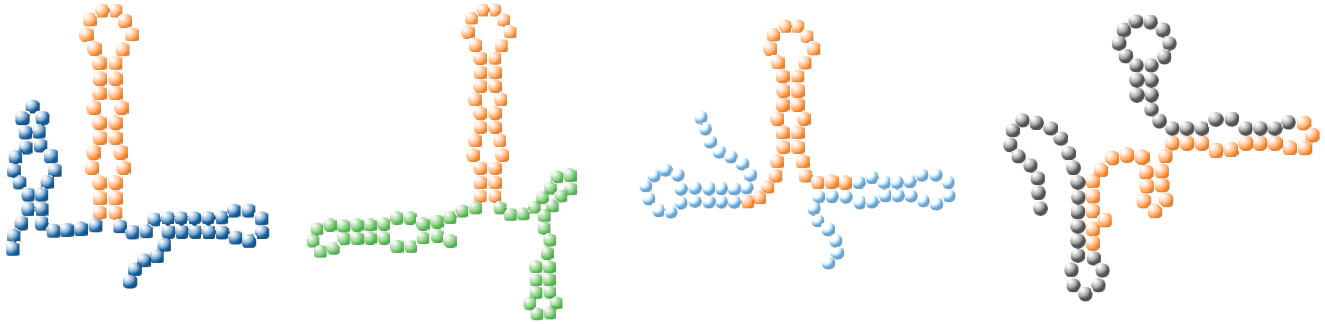
Q1: Modular robustness of RNA

Q1 Does the orange block change after insertion?
Do the blue blocks make a difference in this?





Quantifying RNA structural robustness



GOAL

Measure the intrinsic tendency of an RNA sequence to robustly fold independent of its sequence context, which we define as **self containment**

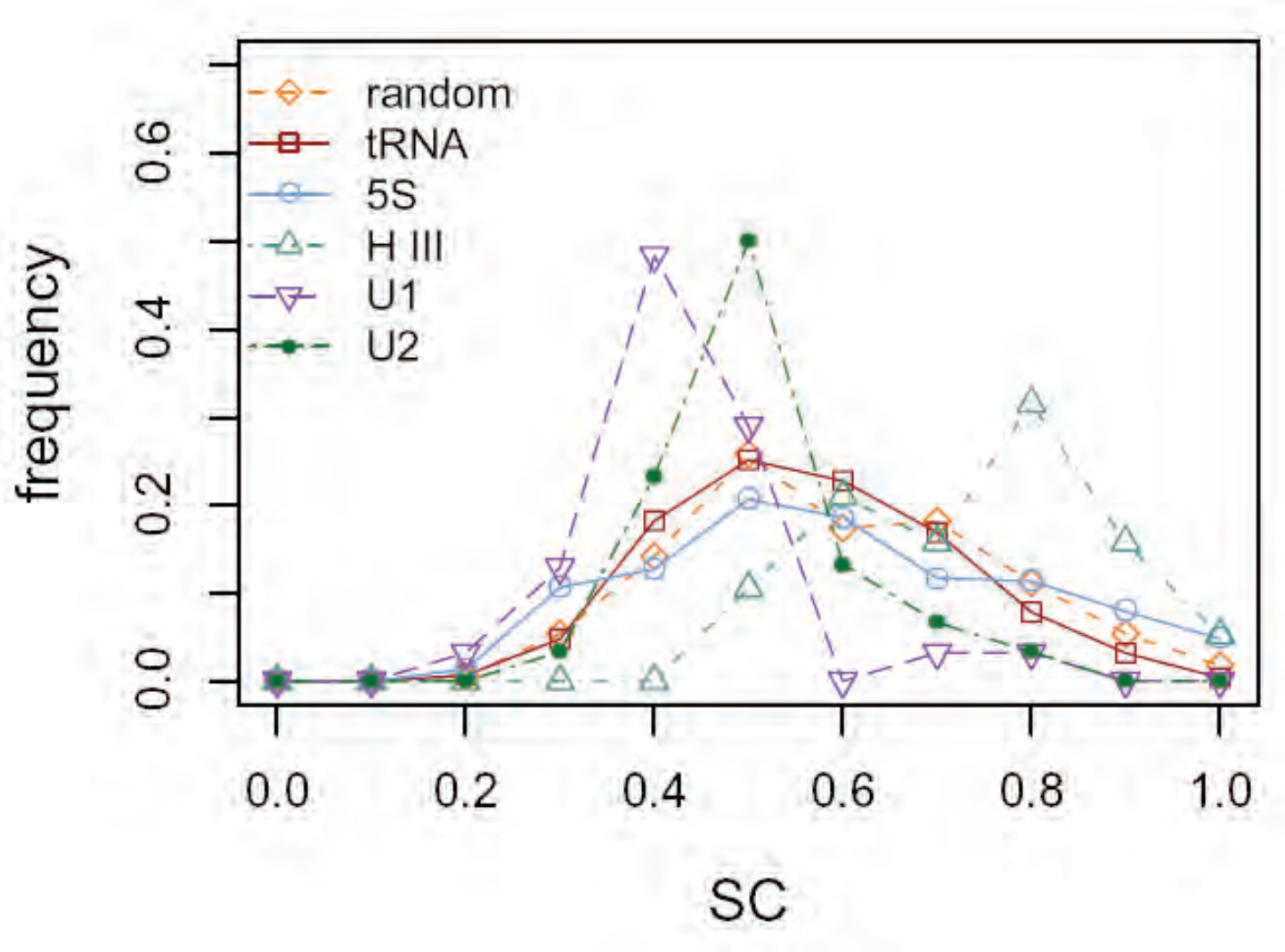
STRATEGY

Given an RNA sequence of interest with structure **S**:

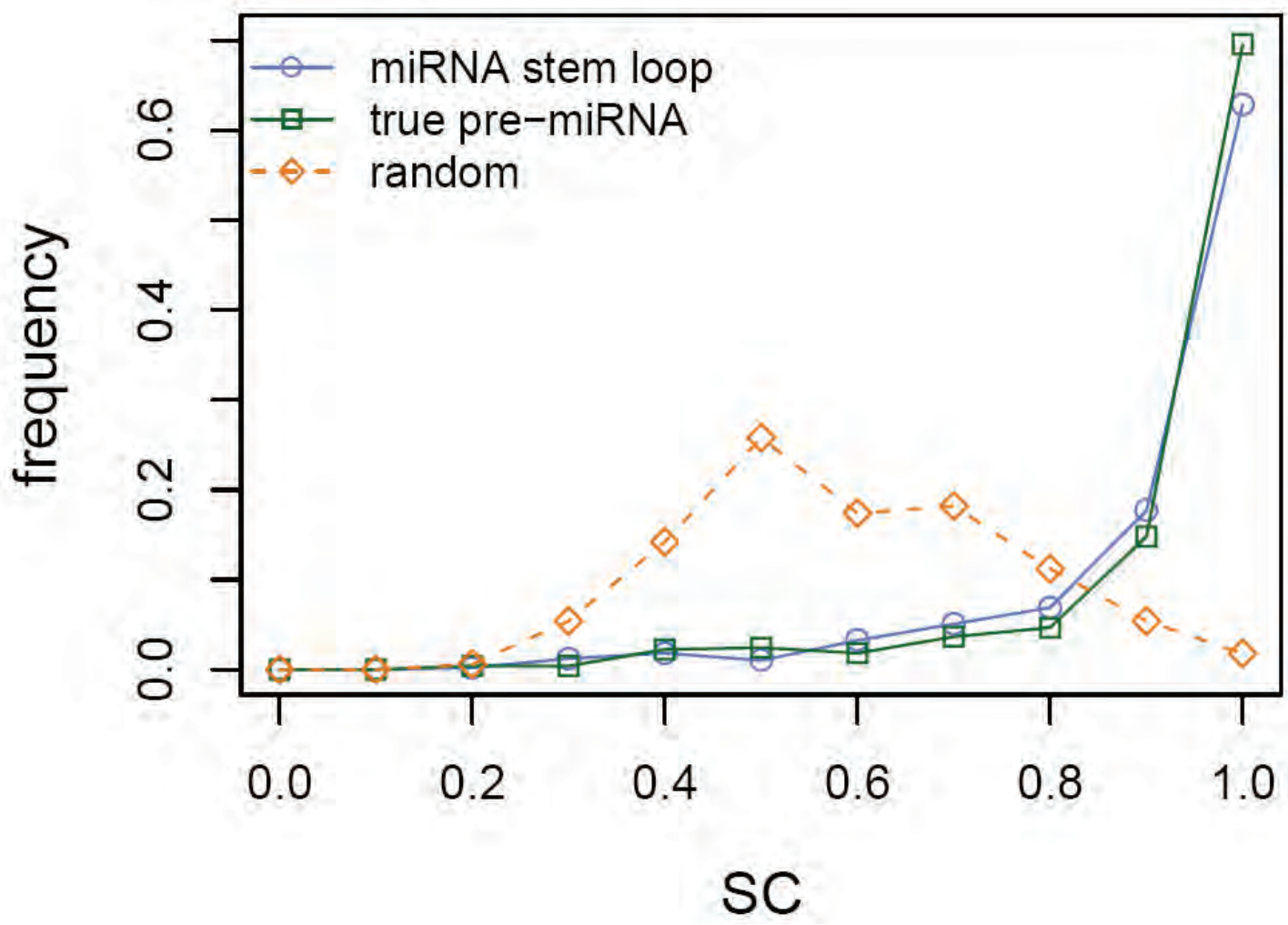
1. Generate a random simulated sequence context
2. Computationally predict the folded structure of the composite RNA, **S'**
3. Measure the proportion of **S** present in **S'**

Repeat *n* times and average the proportions

RNA families are generally not self contained...



...except for microRNAs

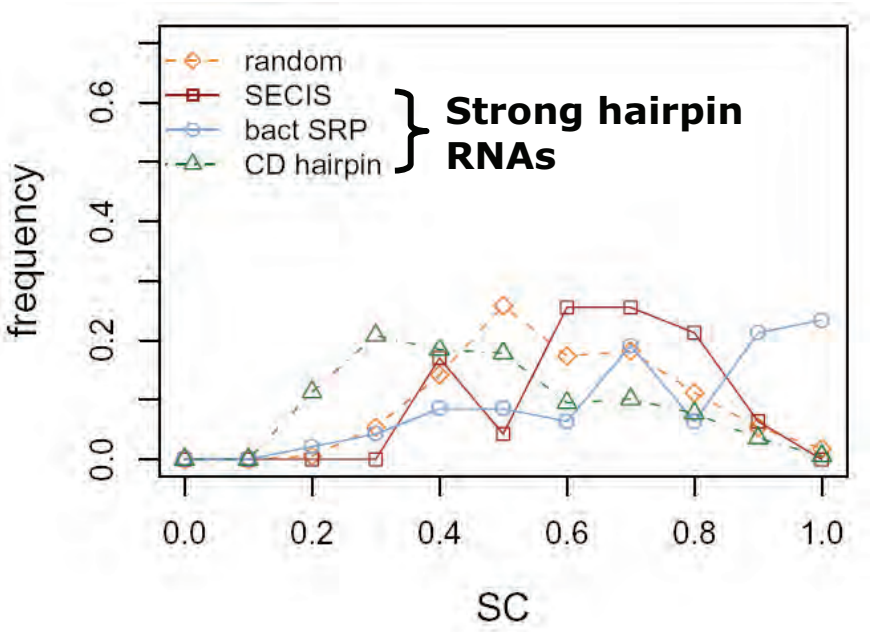


What's special about microRNAs?

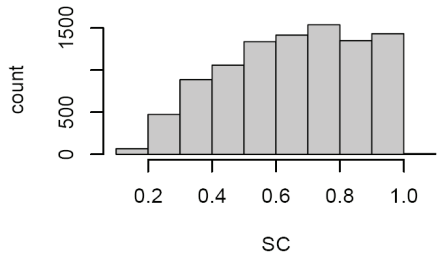
mir-1

miRNAs are modulators of gene expression with an intermediate "**hairpin**" structural form

- Hairpin shape does not imply high self containment



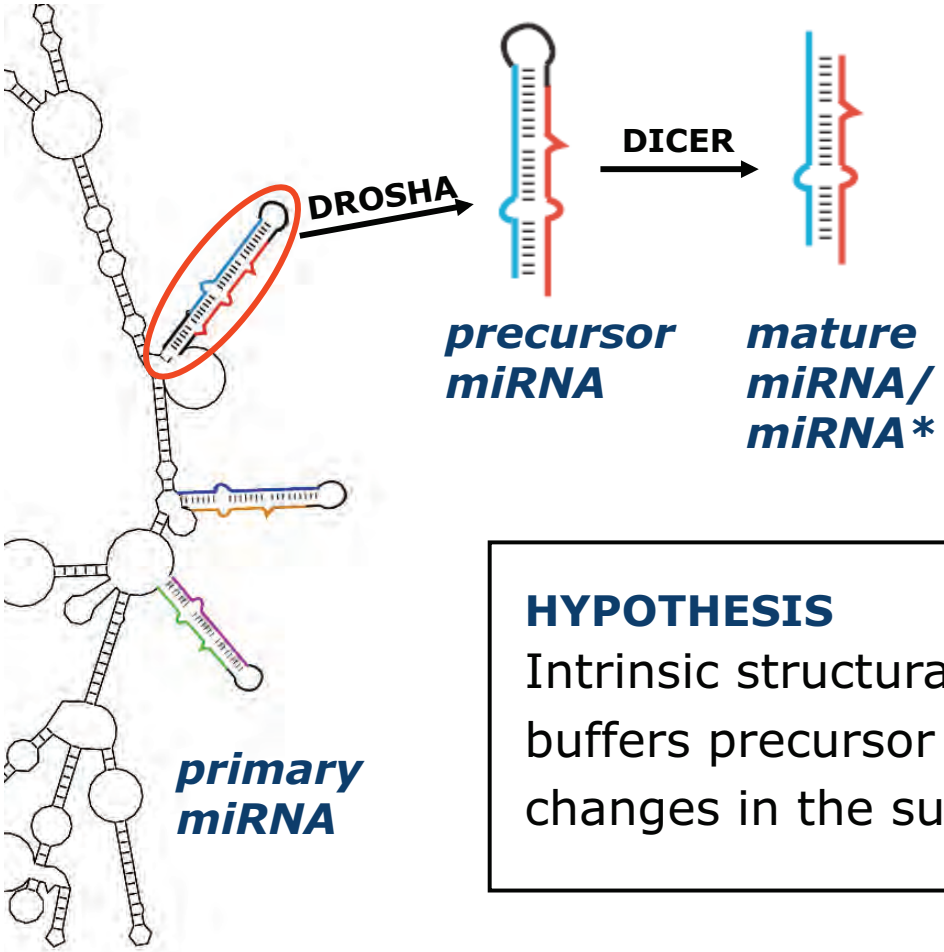
Distribution of all hairpins across all annotated RNA families



Avg SC = 0.65

miRNA biogenesis and genomic organization

- **miRNAs are produced by processing larger precursors**



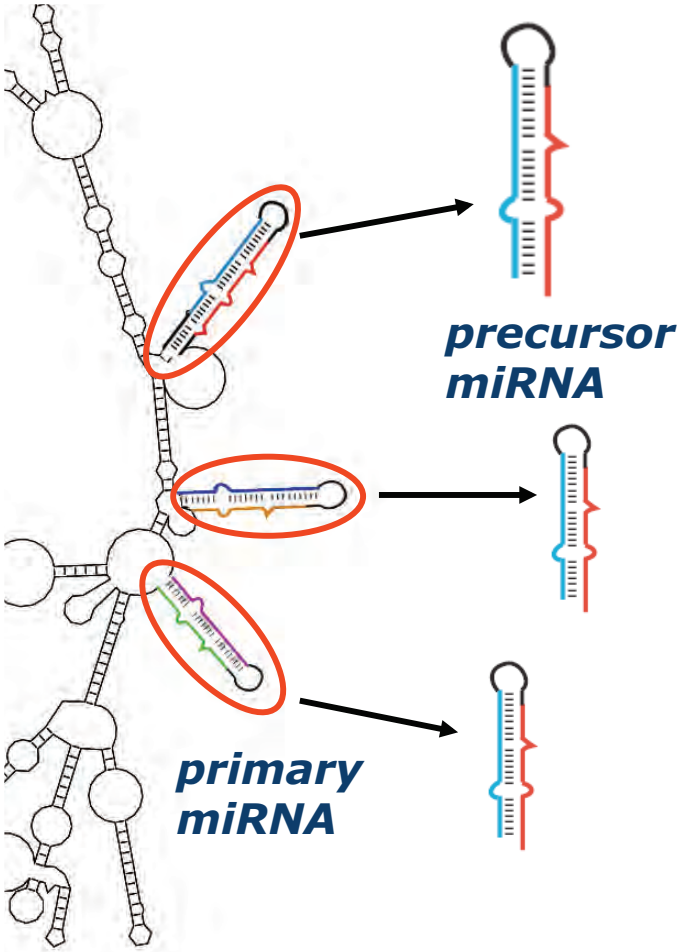
- Processing miRNAs is structurally specific
- Without structural robustness of the pre-miRNA, the processing pathway is disrupted

HYPOTHESIS

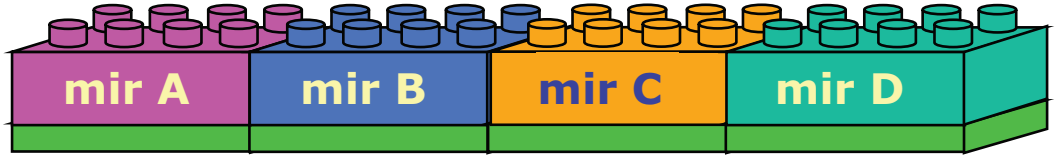
Intrinsic structural robustness, i.e. modularity, buffers precursor miRNAs against sequence changes in the surrounding primary transcript

miRNA biogenesis and genomic organization

▶ **Multiple miRNAs are present in the primary transcript**



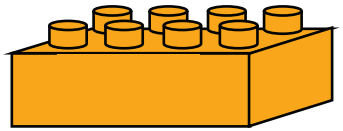
- Primary transcripts can harbor several different miRNAs, each with different functional specificities
- Hypothetical role of primary transcript as an organizing center for miRNAs

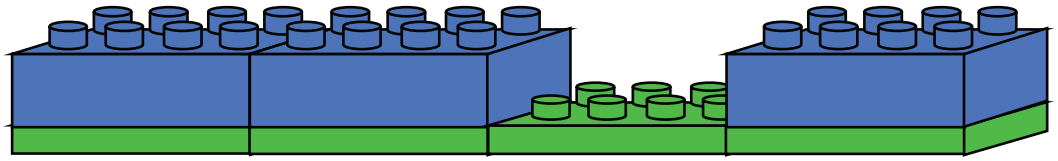


▶ miRNAs occurring in clusters have significantly higher self containment

Q2: Placement of RNA functional modules

Building an RNA molecule...





Composite RNA

Q1 Does the orange block change after insertion?
Do the blue blocks make a difference in this?

Q2 Are the blue blocks changed by the insertion?
Does it even matter if they are changed?

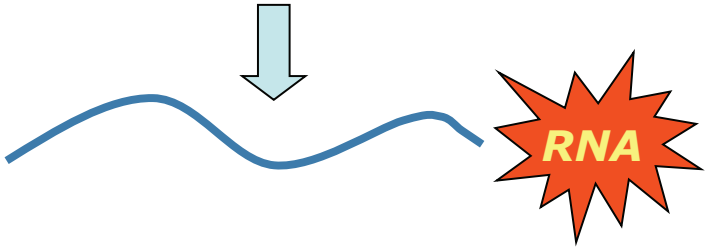
mRNAs can harbor extra signals

GENE (DNA)

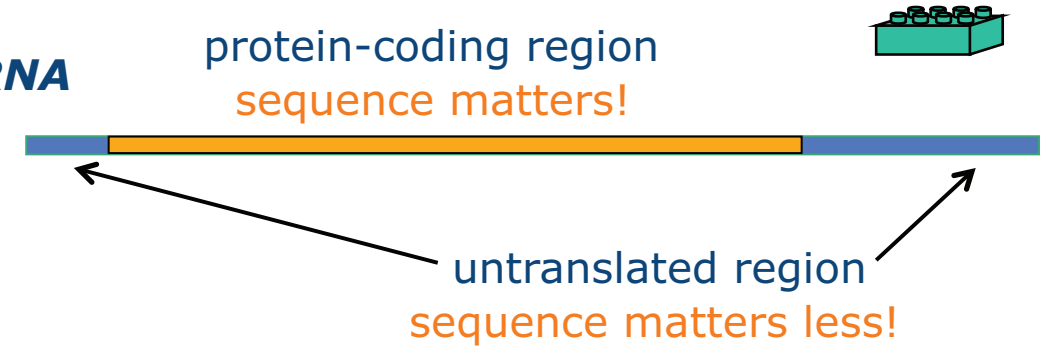
mRNA

messenger RNA

Specific sequence required

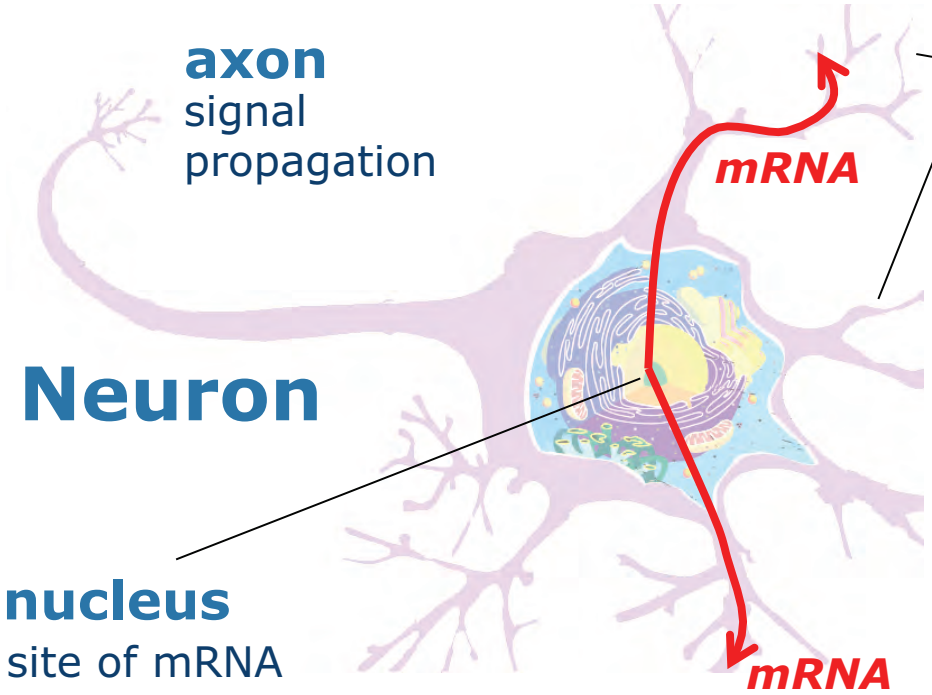


Anatomy of an mRNA



- Regions of the mRNA not necessary for specifying the protein product can contain higher-order information, e.g. for expression regulation

Some mRNAs are moved around the cell



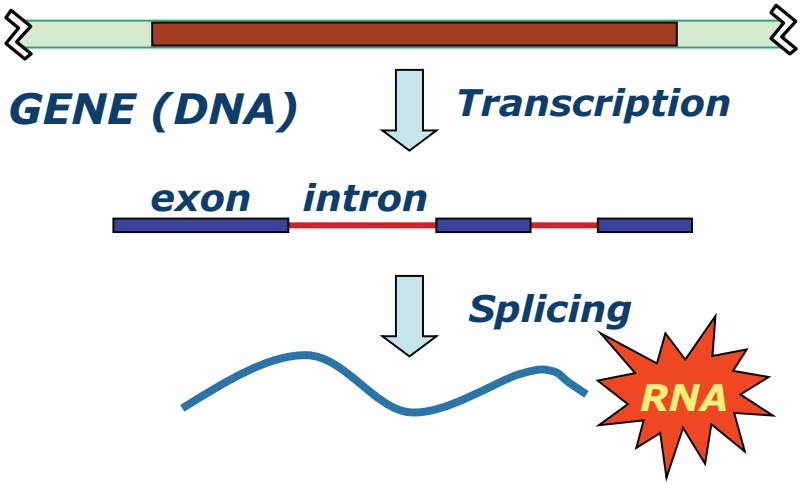
- Structured localization signals can be contained in the untranslated region of an mRNA

e.g. - CamKIIa

▶ Many more mRNAs exhibit localization that don't have obvious signals in the UTR

...maybe the signals are somewhere else

Introns are spliced from mRNAs ...usually

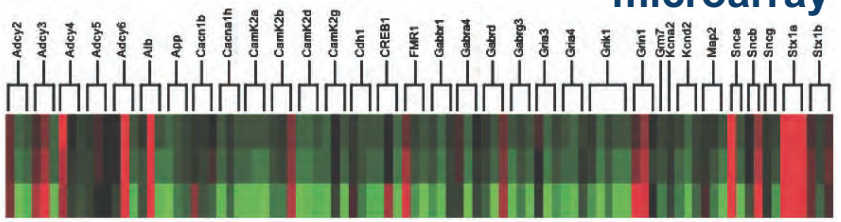


- Introns are parts of the gene that do not contain protein-coding information
- Typically removed to form the mature mRNA before exiting the nucleus

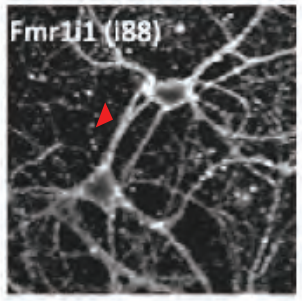
▶ Some introns are not removed in rat neurons

- Multiple experiments indicate the presence of intronic sequence outside the nucleus in rat neurons

microarray

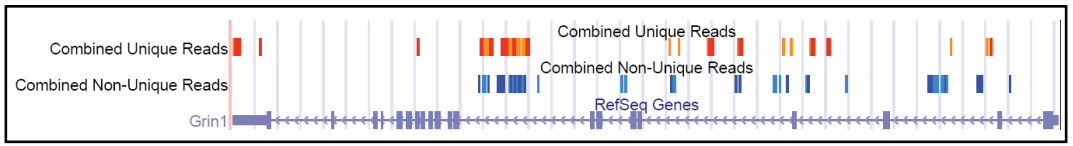


in situ

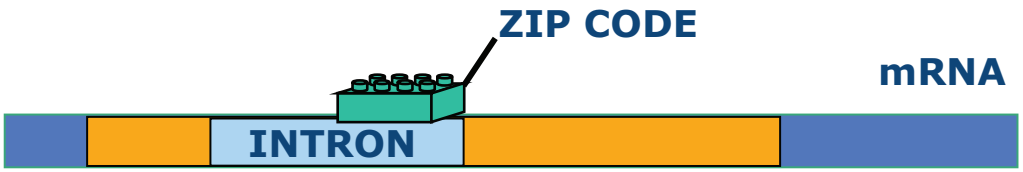


P. Buckley

Solexa sequencing



Localization elements could exist in introns



HYPOTHESIS

Introns that are retained in (quasi-) mature mRNAs can contain localization signals that do not disrupt the protein coding instructions

- ▶ Subsequent removal of the intron occurs once the mRNA has been transported to the right cellular location

What do the signals look like?

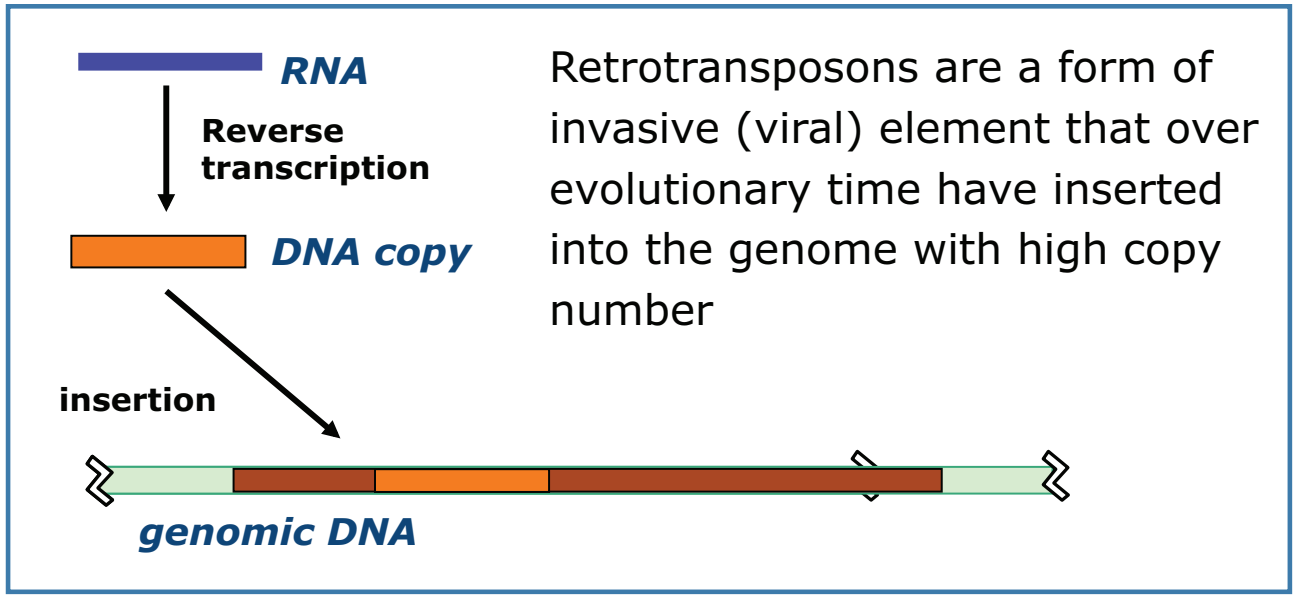
- Typical approach is to search for shared sequence/structure motifs in the sequences of interest (e.g., all retained introns)
- To limit false signals, restrict the search to sequence regions that are **evolutionarily conserved**
  - ▶ Reasoning: if an element is important, it should also exist in related species

Candidate localization element is not conserved

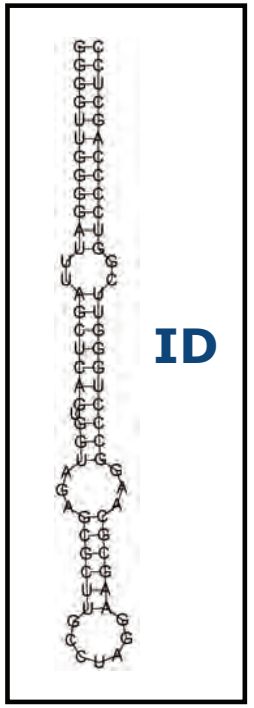
OUR FINDINGS

We did not find any evolutionarily conserved elements...

- We DID find an abundance of one class of "junk" DNA fragments, something typically ignored due to assumed non-functionality: the **ID element**, a type of rodent-specific **SINE retrotransposon**



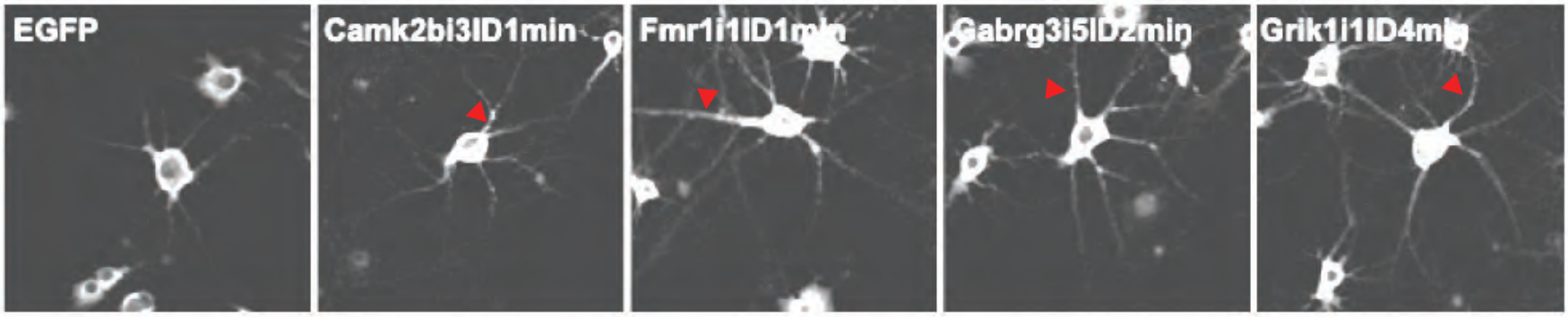
Retrotransposons are a form of invasive (viral) element that over evolutionary time have inserted into the genome with high copy number



ID elements retain a localization signal

ID elements derive from a type of RNA containing a known localization signal: cell body → dendrite

- Do the ID elements retain localization competency? **YES.**



EGFP reporter constructs containing ID elements derived from retained introns are localized to the dendrites

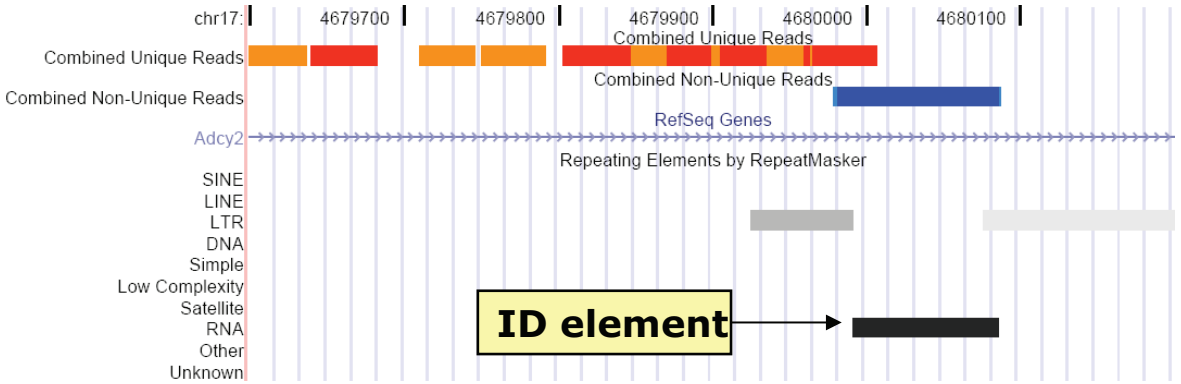
P. Buckley

Transcriptome sequence contains ID elements

- Are the ID elements present in mRNA transcripts? **LIKELY.**

  Solexa short-read sequence data show a significant enrichment in ID-derived sequence among the mRNA population in rat dendrites

  - 63K sequence reads match the ID element sequence
  - 12.5K sequence reads match the B2 SINE element
  - But genome-wide, 90K B2 elements overlap introns compared to only 65K ID elements



Based on neighboring sequence read matches, we identify several intronic ID loci likely to be retained in mRNAs

ID-containing genes have neuronal function

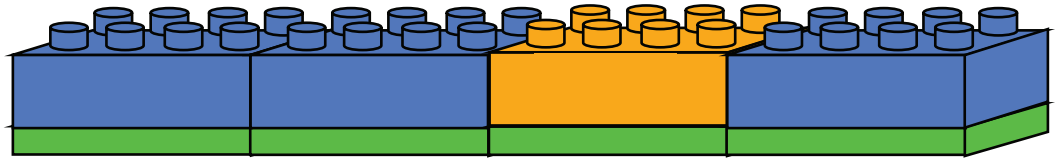
▶ Do genes containing ID elements need to be transported to dendrites? **LIKELY.**

Genes with a surplus of putatively functional ID elements tend to have neuronal function as annotated by the Gene Ontology

- **receptor activity**
- **signal transducer activity**
- **molecular transducer activity**
- **postsynaptic density**
- **system process**
- **synapse**
- **neurological system process**
- **cell-cell signaling**
- **synapse part**
- **neurotransmitter secretion**
- **axon part**
- **generation of signal - cell-cell**
- **reg. of neurotransmitter levels**
- **regulated secretory pathway**
- **synaptic transmission**

In closing: modularity is achieved in different ways

RNAs are sometimes composed of modular components



Composite RNA

Q1 Does the orange block change after insertion?

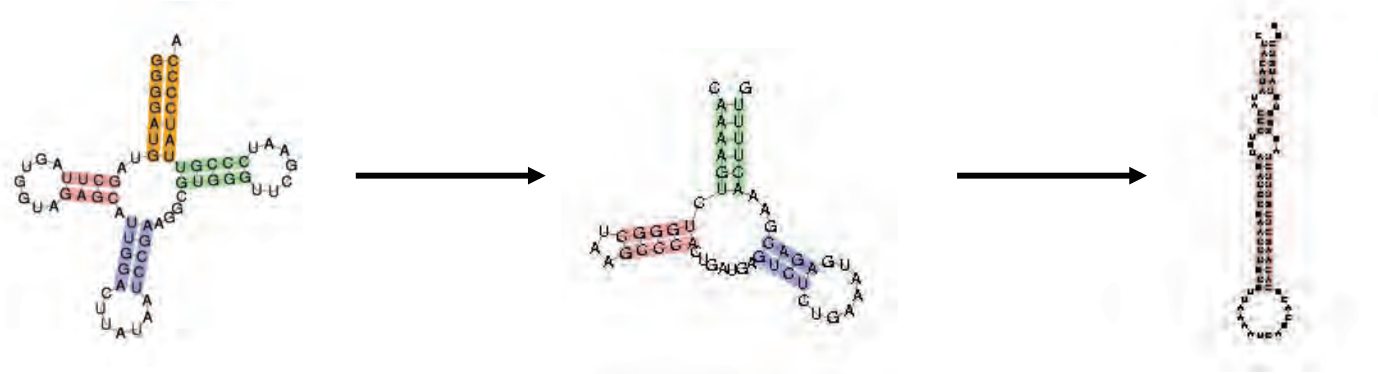
- Some RNAs (microRNAs) are structurally robust, ensuring they don't change in different contexts

Q2 Are the blue blocks changed by the insertion?

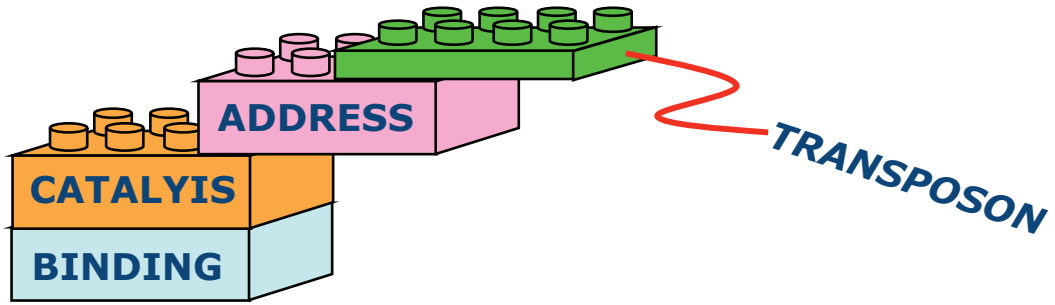
- Structural elements (localization signals) can occur in contexts (introns) where they do not disrupt the function of the containing RNA

Implications: RNA evolution

How did the diversity of RNA function and structure evolve?



Does it make sense to envision modern RNAs as composed of sequence/structural/functional building blocks?



Acknowledgements

- Advisor: Junhyong Kim
- Collaborators: James Eberwine, Peter Buckley
- Kim lab members past and present
- Thesis committee
- Genomics and Computational Biology Program
- U.S. Dept. of Energy, Krell Institute

References:

Lee & Kim, 2008. Self containment, a property of modular RNA structures, distinguishes microRNAs. PLoS Comp Biol 4(8).

Buckley & Lee, et al. Messenger RNA localized to dendrites retain specific intronic sequences. Submitted.

http://kim.bio.upenn.edu/~miler/