

# Long Noncoding RNA in the North American Beaver

Amita Kashyap<sup>1, 4</sup>; Adelaide C Rhodes<sup>2, 4</sup>; and Stephen A Ramsey, PhD<sup>3, 4</sup>

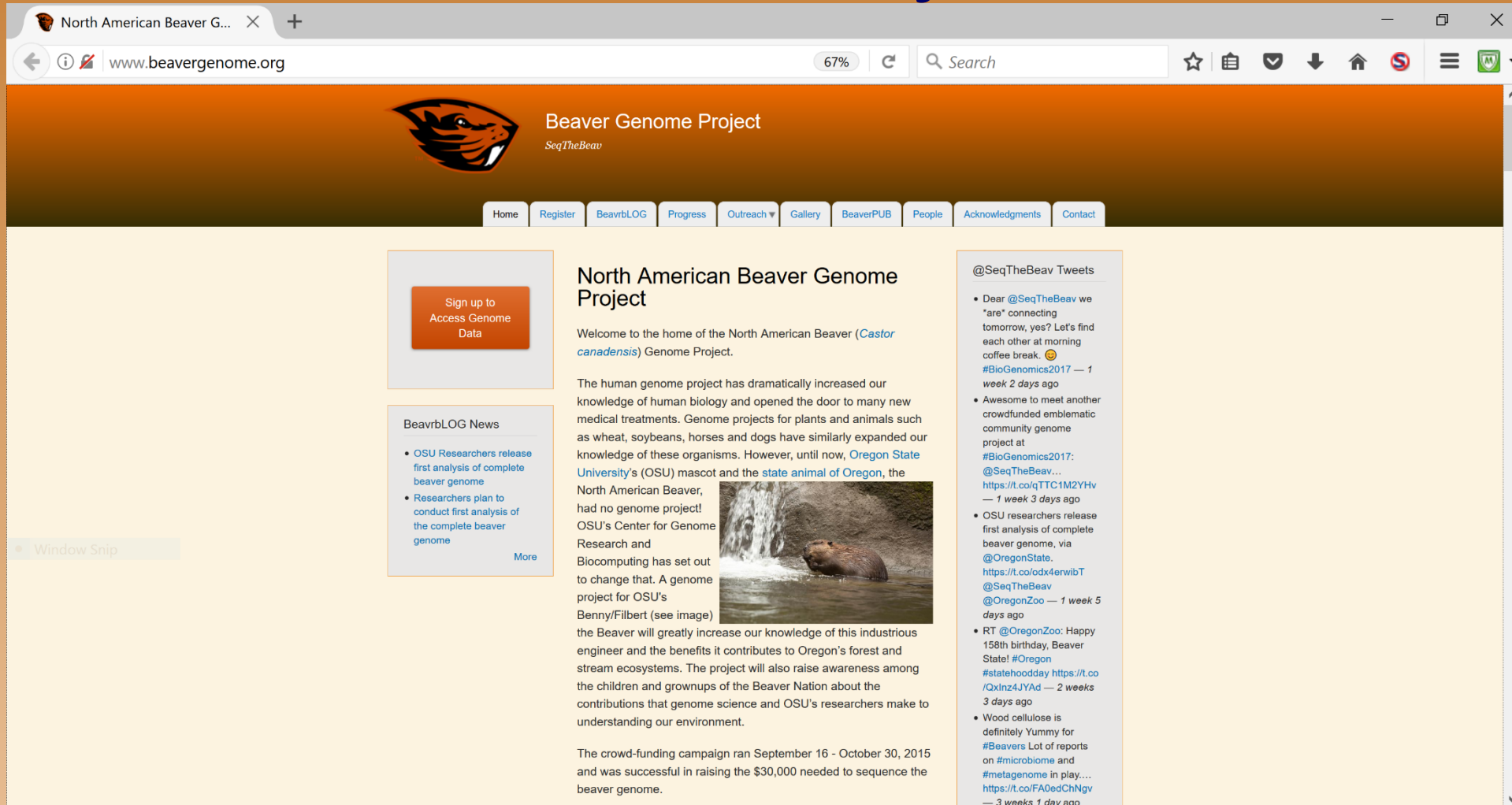
<sup>1</sup>BioResource Research, College of Agricultural Sciences

<sup>2</sup>Center for Genome Research and Biocomputing

<sup>3</sup>Department of Biomedical Sciences

<sup>4</sup>Oregon State University, Corvallis

# This research is part of OSU's Beaver Genome Project



The screenshot shows a web browser window with the address bar displaying "www.bevergenome.org". The website has an orange header with a beaver logo and the text "Beaver Genome Project" and "SeqTheBeav". Below the header is a navigation bar with links: Home, Register, BeaverLOG, Progress, Outreach, Gallery, BeaverPUB, People, Acknowledgments, and Contact. The main content area is divided into three columns. The left column has a "Sign up to Access Genome Data" button and a "BeavrLOG News" section with two bullet points. The middle column has a heading "North American Beaver Genome Project" and a welcome message. Below this is a paragraph about the project's goals and a photo of a beaver. The right column has a "@SeqTheBeav Tweets" section with several tweets. The bottom of the page has a "Window Snip" button.

North American Beaver G... X +

www.bevergenome.org 67% Search

Beaver Genome Project  
SeqTheBeav

Home Register BeaverLOG Progress Outreach Gallery BeaverPUB People Acknowledgments Contact

Sign up to Access Genome Data

BeavrLOG News

- OSU Researchers release first analysis of complete beaver genome
- Researchers plan to conduct first analysis of the complete beaver genome

More

## North American Beaver Genome Project

Welcome to the home of the North American Beaver (*Castor canadensis*) Genome Project.

The human genome project has dramatically increased our knowledge of human biology and opened the door to many new medical treatments. Genome projects for plants and animals such as wheat, soybeans, horses and dogs have similarly expanded our knowledge of these organisms. However, until now, [Oregon State University's](#) (OSU) mascot and the [state animal of Oregon](#), the North American Beaver, had no genome project! OSU's Center for Genome Research and Biocomputing has set out to change that. A genome project for OSU's Benny/Filbert (see image) the Beaver will greatly increase our knowledge of this industrious engineer and the benefits it contributes to Oregon's forest and stream ecosystems. The project will also raise awareness among the children and grownups of the Beaver Nation about the contributions that genome science and OSU's researchers make to understanding our environment.

The crowd-funding campaign ran September 16 - October 30, 2015 and was successful in raising the \$30,000 needed to sequence the beaver genome.

@SeqTheBeav Tweets

- Dear @SeqTheBeav we "are" connecting tomorrow, yes? Let's find each other at morning coffee break. ☺  
#BioGenomics2017 — 1 week 2 days ago
- Awesome to meet another crowdfunded emblematic community genome project at #BioGenomics2017: @SeqTheBeav...  
<https://t.co/qTTC1M2YHv> — 1 week 3 days ago
- OSU researchers release first analysis of complete beaver genome, via @OregonState, <https://t.co/odx4erwibT> @SeqTheBeav @OregonZoo — 1 week 5 days ago
- RT @OregonZoo: Happy 158th birthday, Beaver State! #Oregon #statehoodday <https://t.co/Qxinz4JYAd> — 2 weeks 3 days ago
- Wood cellulose is definitely Yummy for #Beavers Lot of reports on #microbiome and #metagenome in play...  
<https://t.co/FA0edChNgv> — 3 weeks 1 day ago

# Overview

- Background
- Objective and Impact
- Methodology
- Results
- Further Research

# Overview

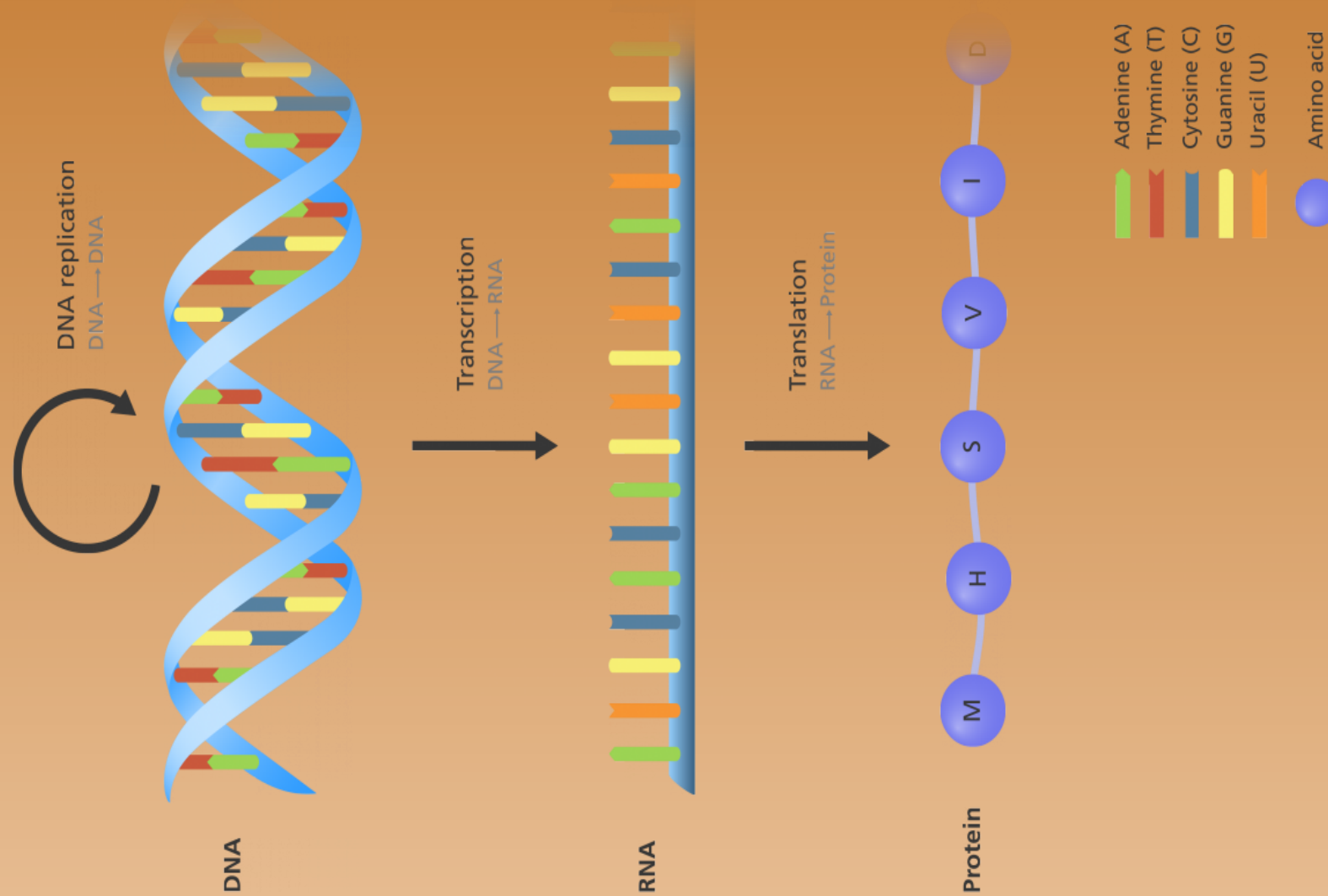
- **Background**
- Objective and Impact
- Methodology
- Results
- Further Research



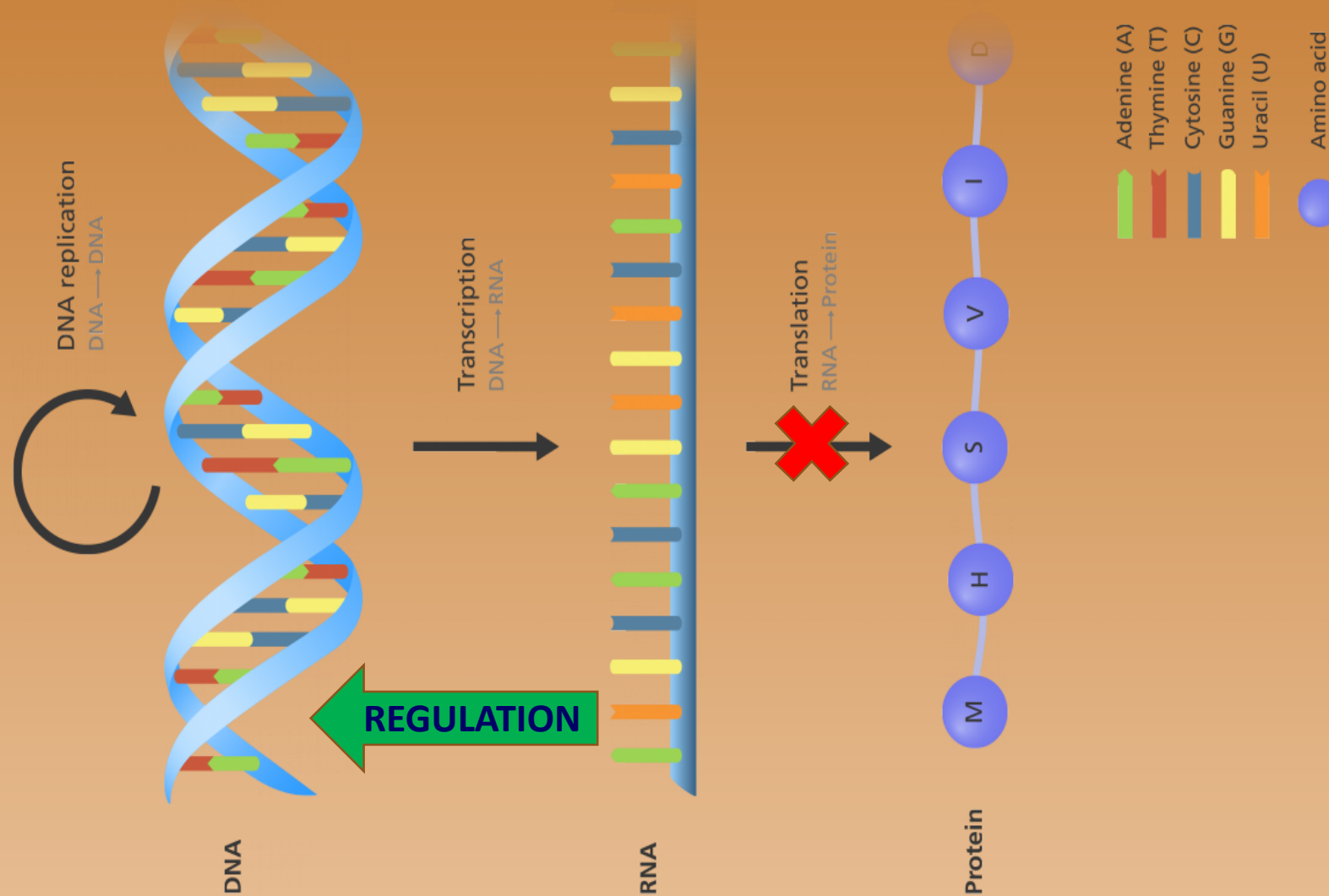
# The beaver has many unique characteristics



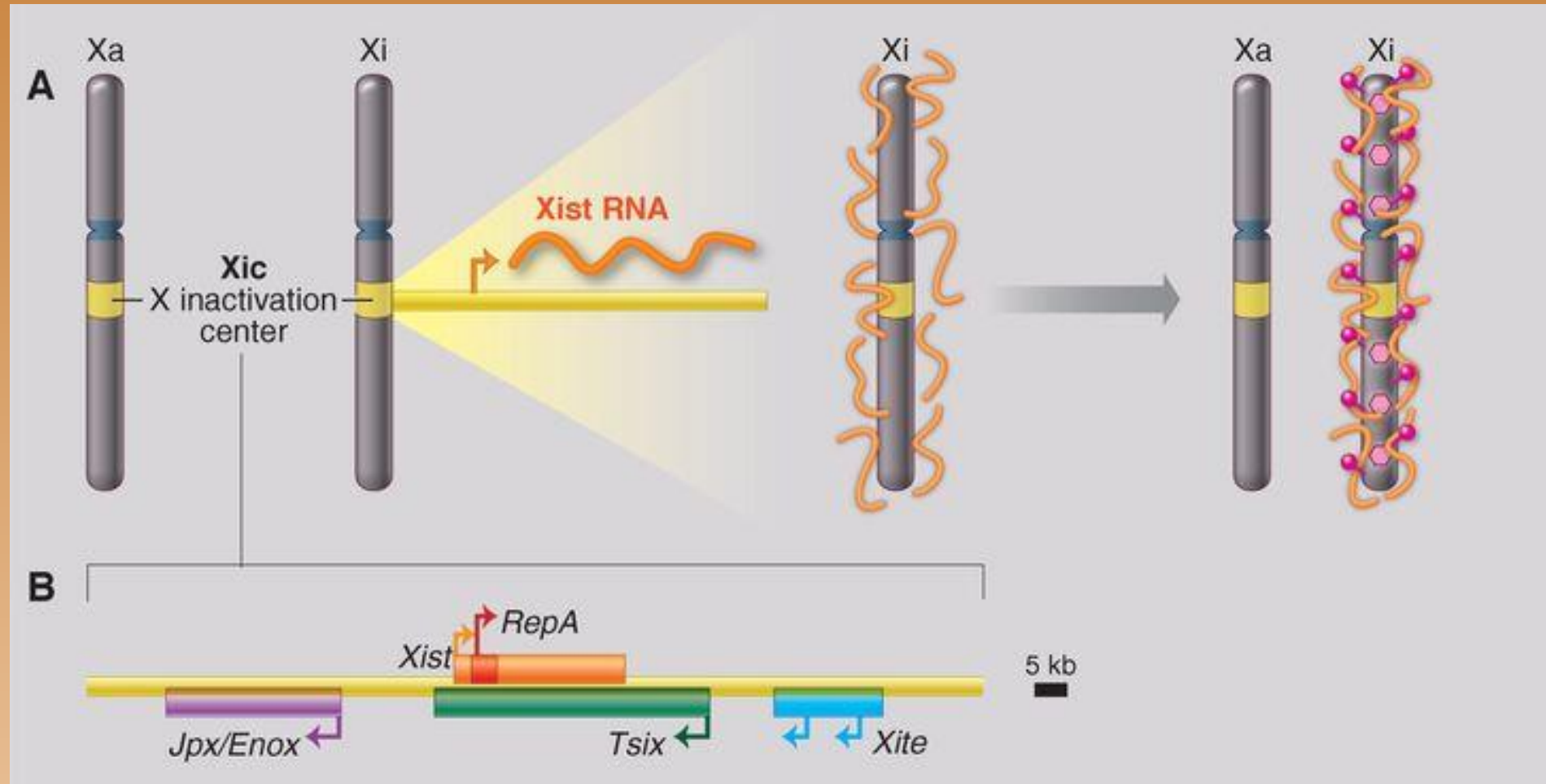
# The Central Dogma is not the whole story



# The Central Dogma is **not** the whole story



# Long noncoding RNA (lncRNA) is regulatory





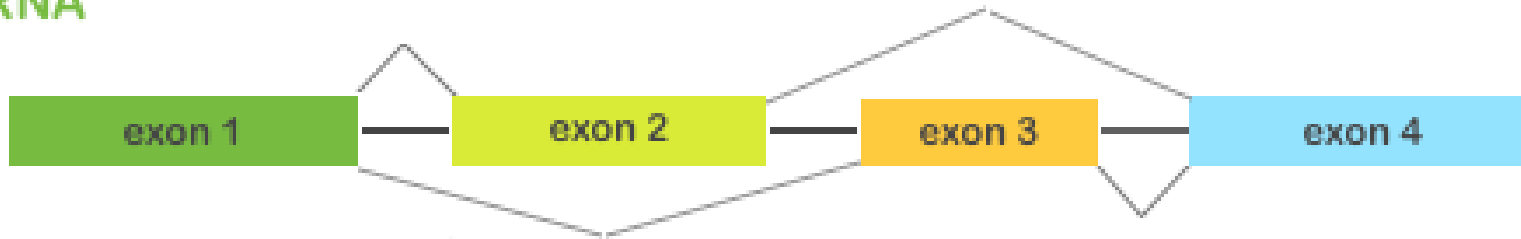
# Known Functions of Long Noncoding RNA

- X chromosome inactivation (mammals)
- Hypoxia response (human, therians)
- DNA damage (human, mouse)
- Osmotic regulation (rat)
- Endoplasmic Reticulum stress (human, mouse)
- Nitric Oxide pathway (mouse)
- Starvation and growth arrest/apoptosis (human, mouse, rat)
- Tumor suppressor/oncogene (therians)
- Viral Infection (placental mammals)
- Microbial Infection (mouse)
- Oxidative Stress (mouse)

Amaral, P. P., Dinger, M. E. & Mattick, J. S. Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective. *Briefings in Functional Genomics* **12**, 254–278 (2013).

# The transcriptome is the sum total of RNA

pre-mRNA



alternative splicing

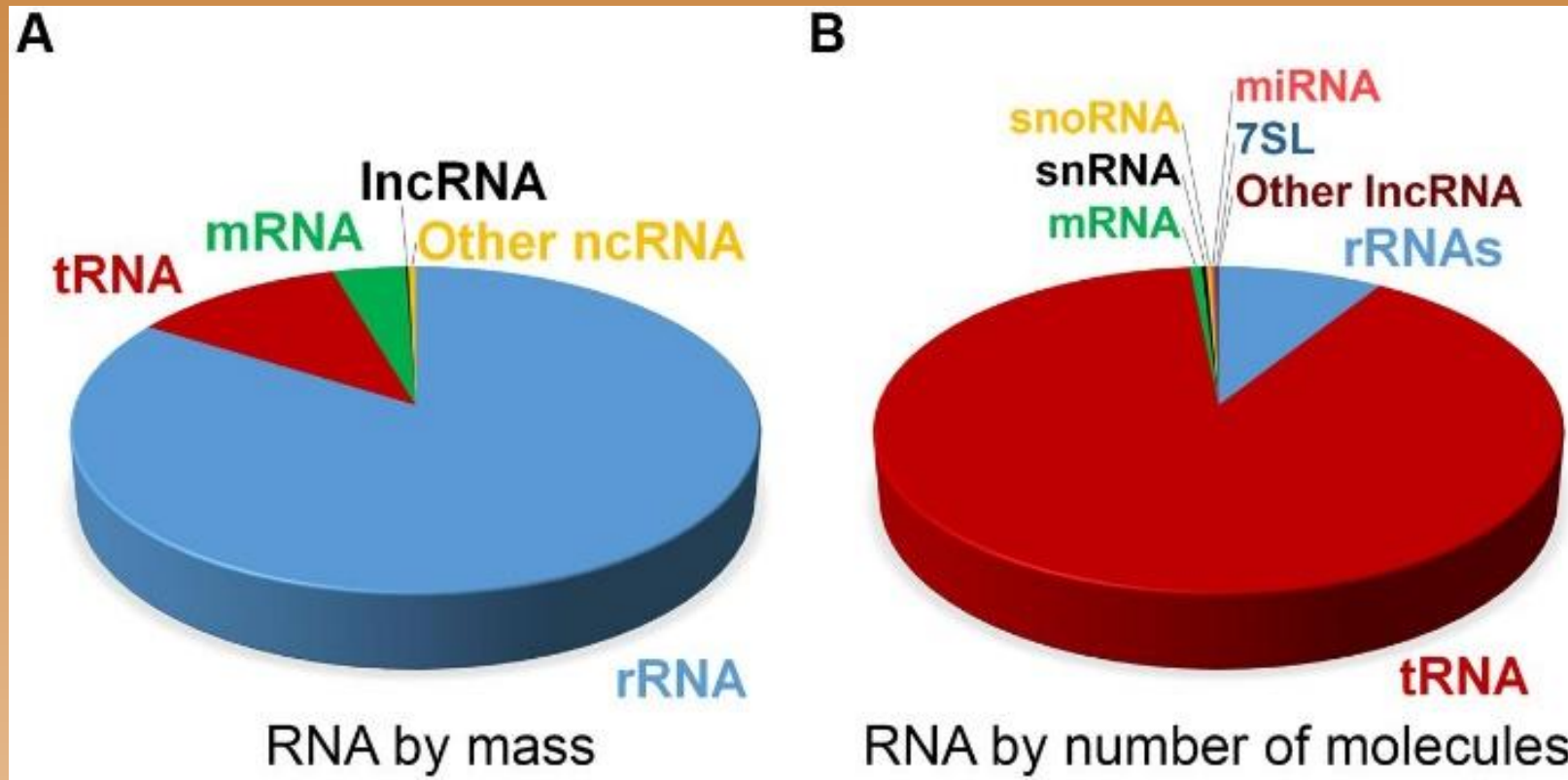


mRNA isoform 1



mRNA isoform 2

# The transcriptome is the sum total of RNA



Palazzo, A. F. & Lee, E. S. Non-coding RNA: what is functional and what is junk? *Frontiers in Genetics* 6, 2 (2015).

# The transcriptome varies by tissue

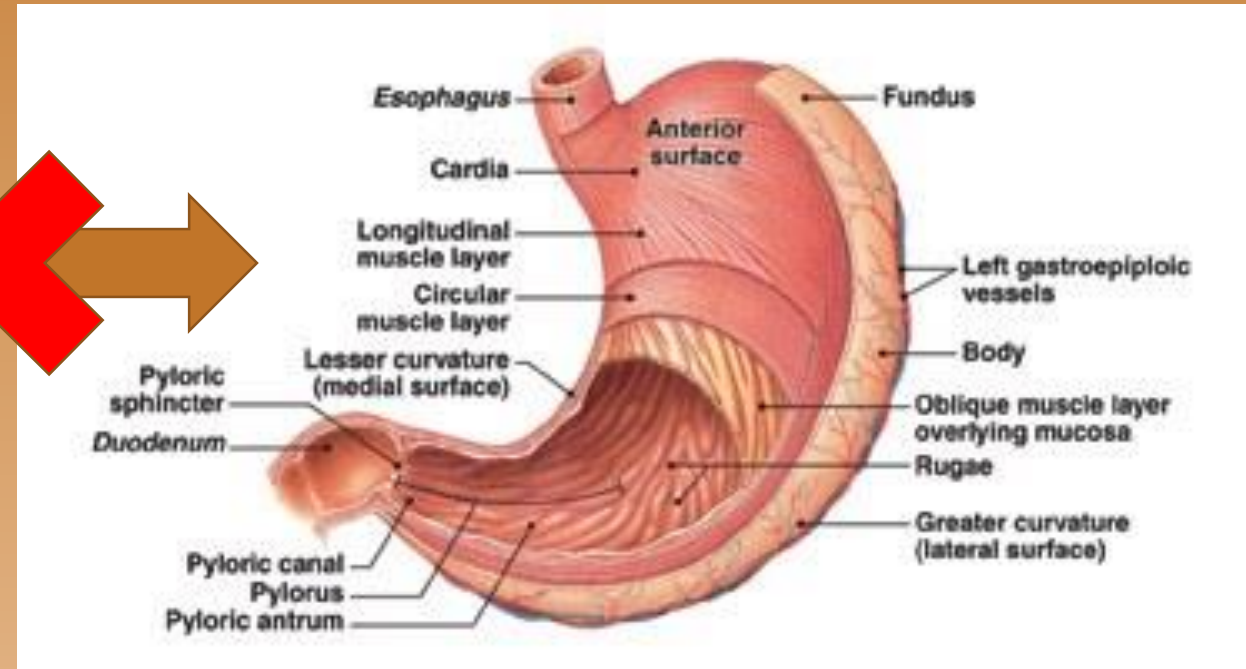
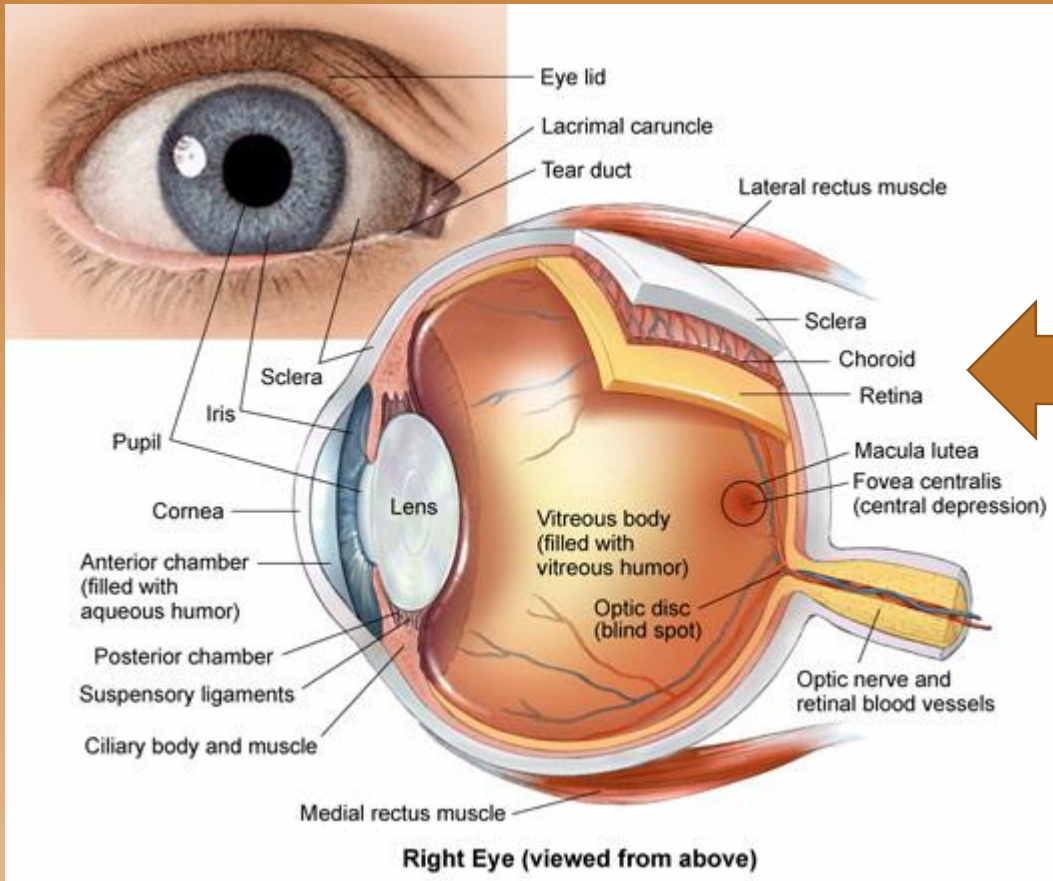


Image Credit: OrgansOfTheBody.com. *Stomach Anatomy*  
Image Credit: Your Eyes Center. *Eye Anatomy*



# Overview

- Background
- **Objective and Impact**
- Methodology
- Results
- Further Research

# Objective

Discover (putative) novel long noncoding RNA in the North American Beaver (*Castor canadensis*)

# Impact of Research

- Shed light on novel characteristics of the beaver
- Provide targets for experimental validation and further research
- Provide insight into lncRNA, applicable across species

# Overview

- Background
- Objective and Impact
- **Methodology**
- Results
- Further Research



# RNA/DNA came from two beavers



# RNA came from 16 tissues

- Heart
- Liver
- Lung
- Intestine
- Ovaries
- Castor gland
- Muscle
- Stomach
- Kidney
- Brain
- Toe webbing
- Spleen
- Blood
- Placenta
- Tail
- Tongue

# Assemblers assemble by overlap

ACTCTCGGGCGTG TAGTAATGC

TAGTAATGC GGGTCTCGAAGTC

# Assemblers assemble by overlap

ACTCTCGGGCGTG**TAGTAATGC**

**TAGTAATGC**GGGTCTCGAAGTC



ACTCTCGGGCGTG**TAGTAATGC**GGGTCTCGAAGTC



# Assemblers assemble by overlap

ACTCTCGGGCGTG**TAGTAATGC**

**TAGTAATGC**GGGTCTCGAAGTC

ACTCTCGGGCGTG**TAGTAATGC**  
**TAGTAATGC**GGGTCTCGAAGTC

ACTCTCGGGCGTG**TAGTAATGC**GGGTCTCGAAGTC

# Assemblers assemble by overlap

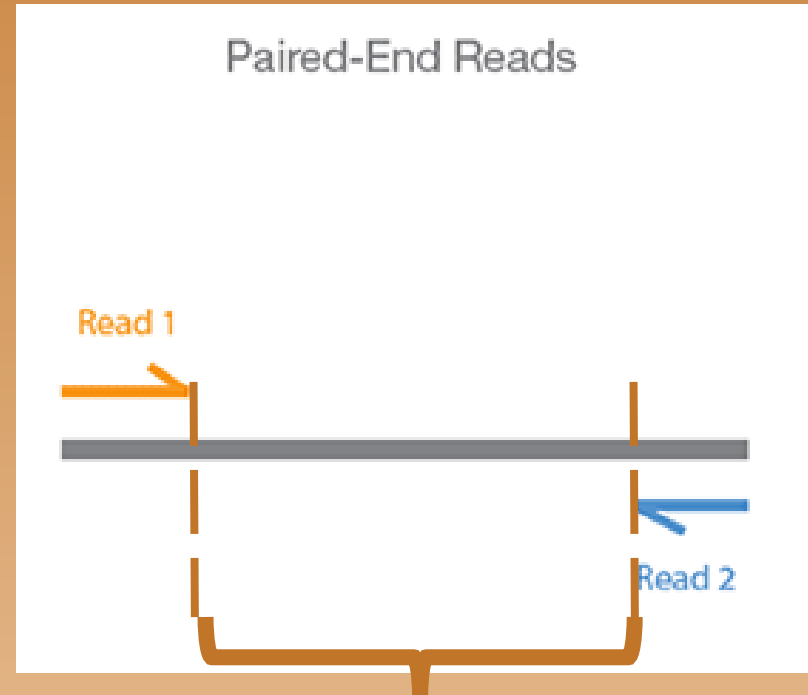
ACTCTCGGGCGTG**TAGTAATGC**

**TAGTAATGC**GGGTCTCGAAGTC

ACTCTCGGGCGTG**TAGTAATGC**  
**TAGTAATGC**GGGTCTCGAAGTC

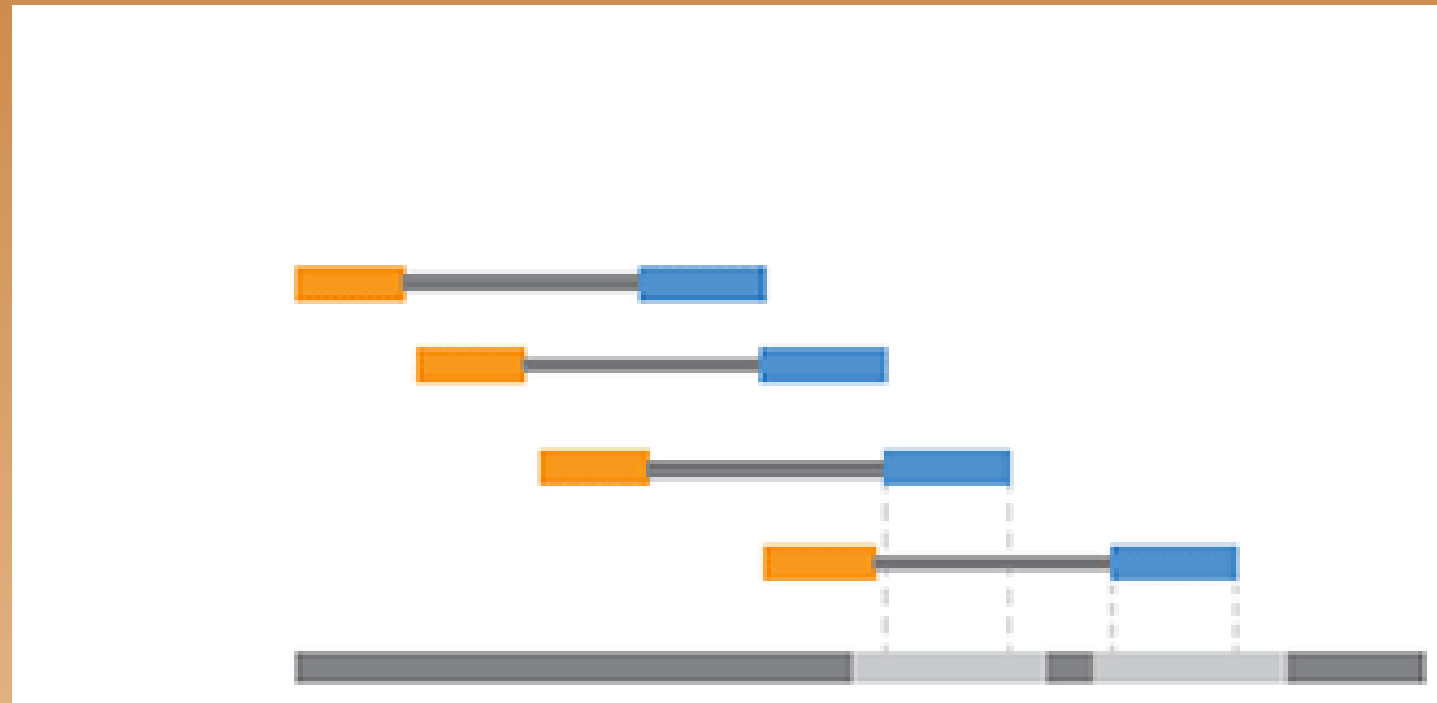
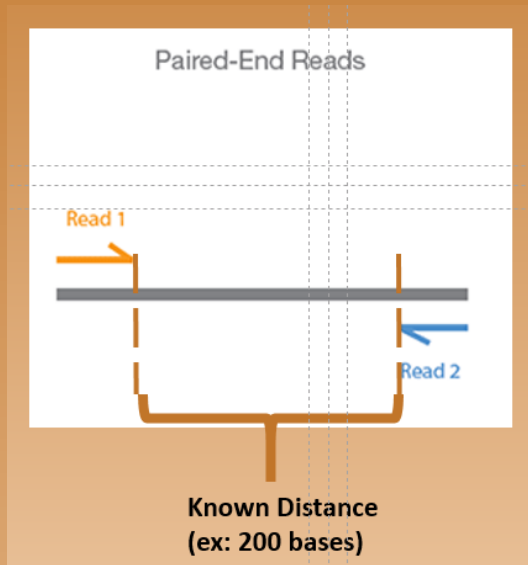
ACTCTCGGGCGTG**TAGTAATGC**GGGTCTCGAAGTC

# Paired-End reads are a known distance from each other



**Known Distance  
(ex: 200 bases)**

# Assemblers take advantage of paired-end reads



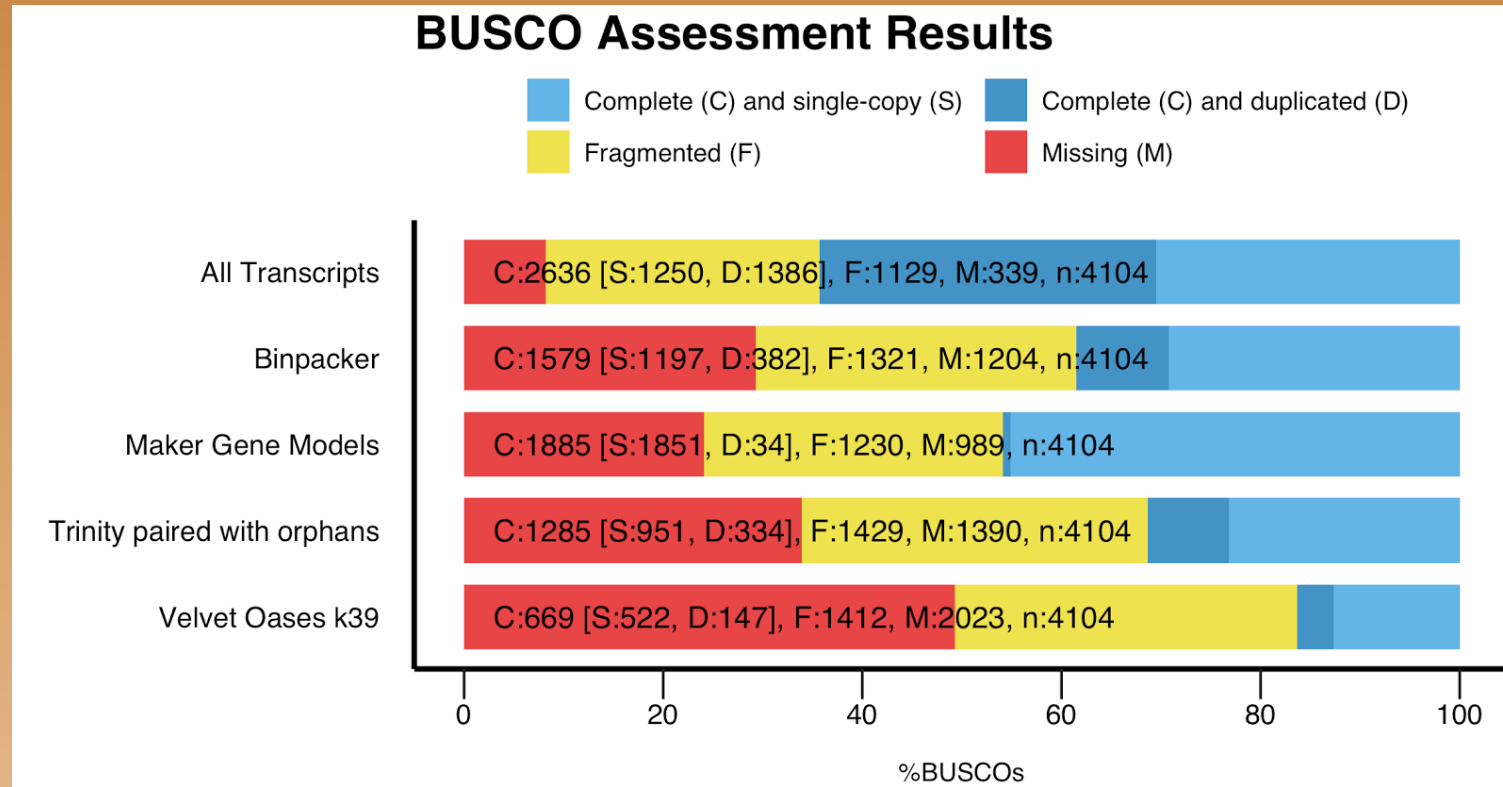


# Many programs were used to construct transcriptome assemblies

- Trinity
  - *de novo*
  - Reference Guided
- Velvet-Oases
- MAKER
- BinPacker



# Our assemblies complement each other



# Merged Transcriptome

# Merged Transcriptome

86714 contigs

# Merged Transcriptome

86714 contigs

**Novel  
lncRNA**

# Merged Transcriptome

86714 contigs

BLASTn

- Known lncRNA
- Other Known Transcripts

- 62.7% eliminated
- 32312 contigs left

## Novel lncRNA

# Merged Transcriptome

86714 contigs

- Known lncRNA
- Other Known Transcripts

- 62.7% eliminated
- 32312 contigs left

BLASTn

(FDR < 0.01)

Not Putative lncRNA  
(Coding Potential Too High)

- 99.4% eliminated
- 182 contigs left

CPAT

Novel  
lncRNA



# Merged Transcriptome

86714 contigs

BLASTn

- Known lncRNA
- Other Known Transcripts

- 62.7% eliminated
- 32312 contigs left

(FDR < 0.01)

Not Putative lncRNA  
(Coding Potential Too High)

- 99.4% eliminated
- 182 contigs left

CPAT

Pfam

Not Putative lncRNA  
(Has Known Protein Domain)

- 0% eliminated
- 182 contigs left

Novel  
lncRNA

# Merged Transcriptome

86714 contigs

- Known lncRNA
- Other Known Transcripts

- 62.7% eliminated
- 32312 contigs left

BLASTn

(FDR < 0.01)

Not Putative lncRNA  
(Coding Potential Too High)

- 99.4% eliminated
- 182 contigs left

CPAT

Pfam

Not Putative lncRNA  
(Has Known Protein Domain)

- 0% eliminated
- 182 contigs left

## Novel lncRNA

# Merged Transcriptome

86714 contigs

BLASTn

- Known lncRNA
- Other Known Transcripts

- 62.7% eliminated
- 32312 contigs left

(FDR < 0.01)

Not Putative lncRNA  
(Coding Potential Too High)

- 99.4% eliminated
- 182 contigs left

CPAT

Pfam

Not Putative lncRNA  
(Has Known Protein Domain)

- 0% eliminated
- 182 contigs left

## Novel lncRNA

# Merged Transcriptome

86714 contigs

BLASTn

- **Known lncRNA**
- Other Known Transcripts

- 62.7% eliminated
- 32312 contigs left

(FDR < 0.01)

Not Putative lncRNA  
(Coding Potential Too High)

- 99.4% eliminated
- 182 contigs left

CPAT

Pfam

Not Putative lncRNA  
(Has Known Protein Domain)

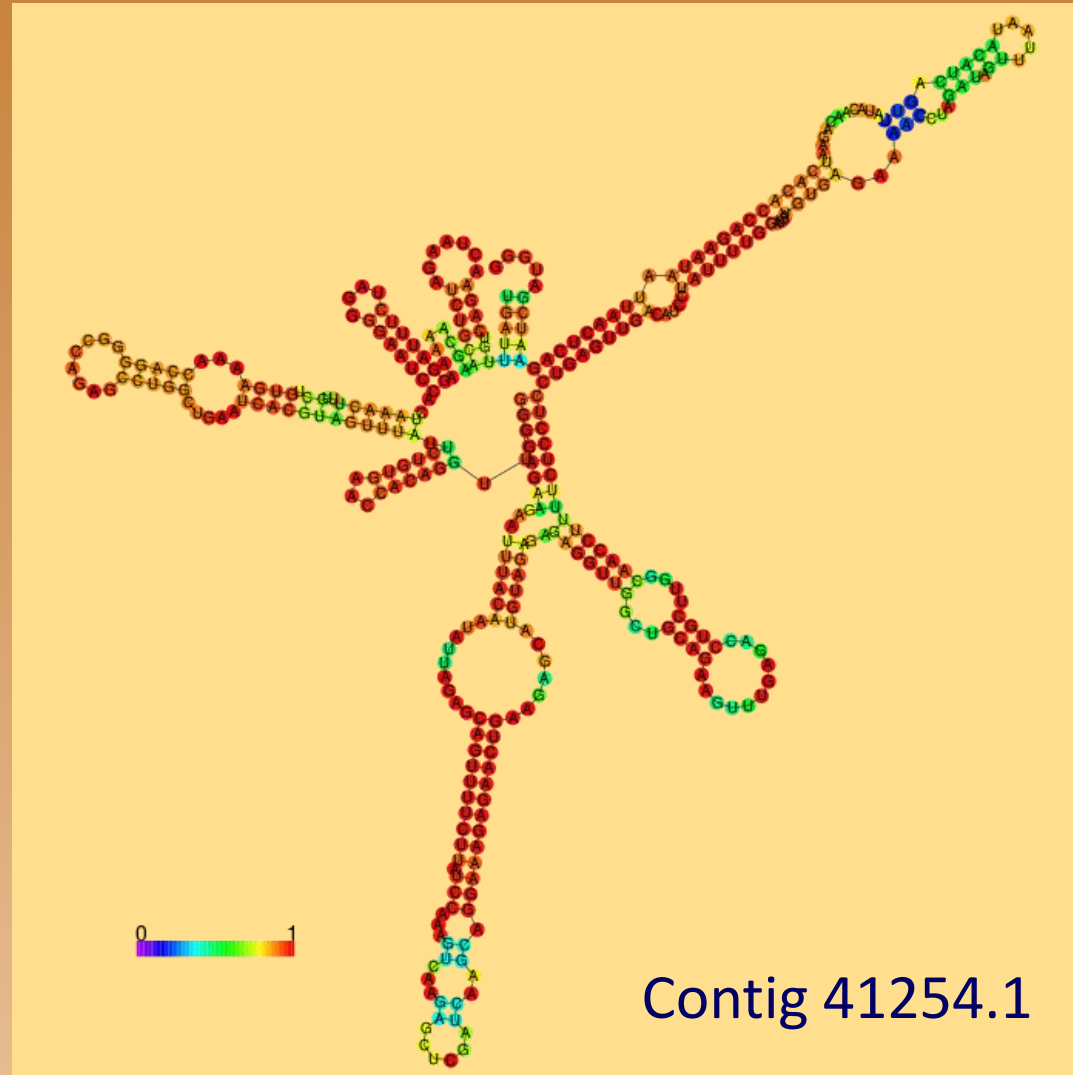
- 0% eliminated
- 182 contigs left

**Novel lncRNA**

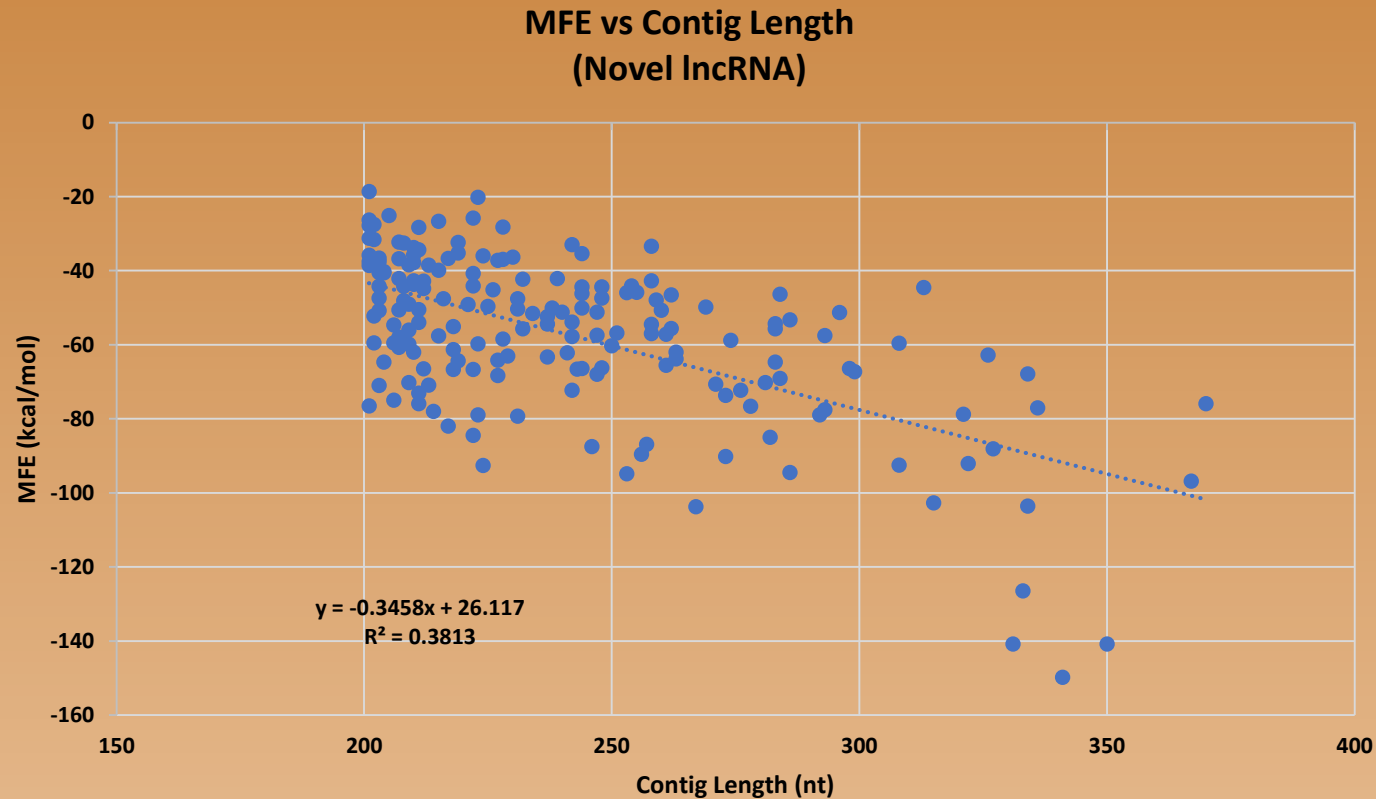
# Overview

- Background
- Objective and Impact
- Methodology
- **Results**
- Further Research

# Secondary Structure Analysis



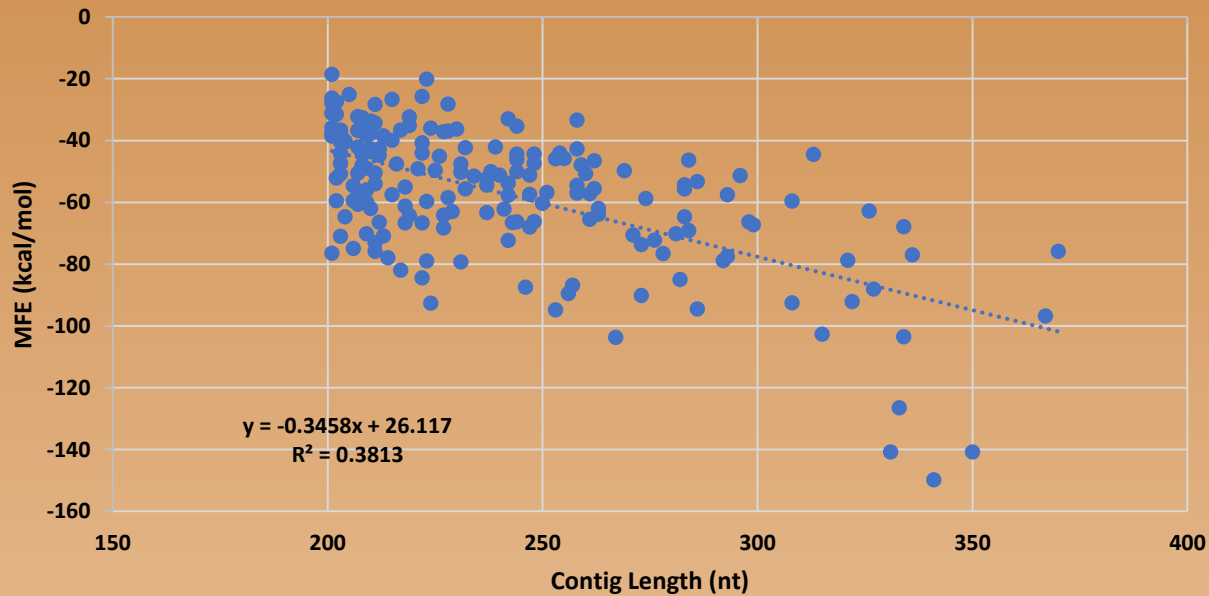
# The novel lncRNA have the expected relationship between length and predicted stability



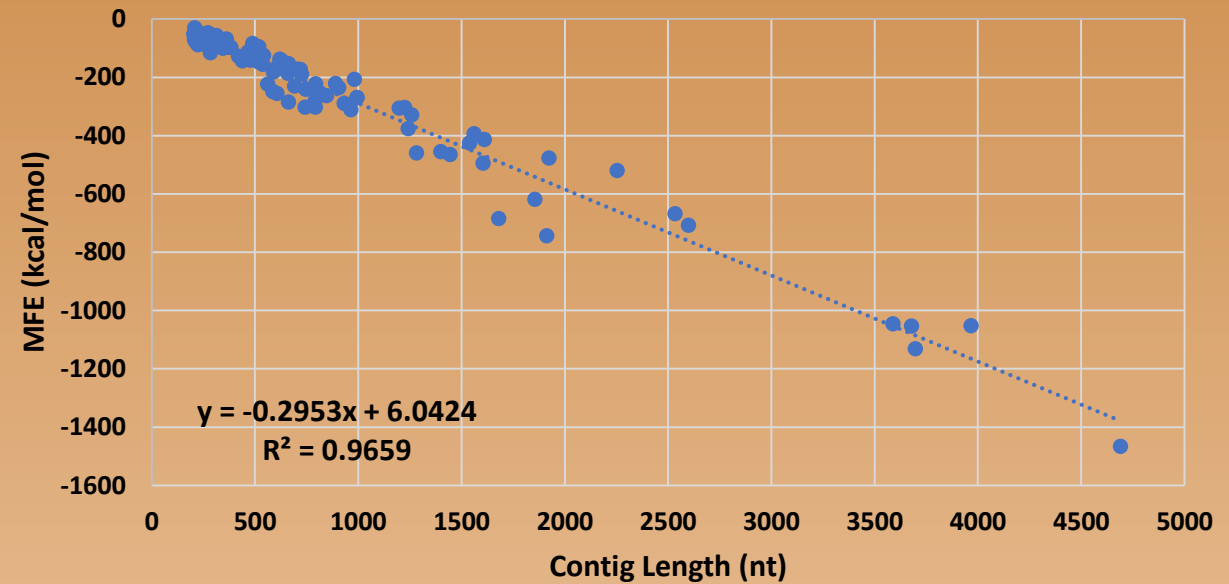


# The novel lncRNA have similar characteristics to the known lncRNA

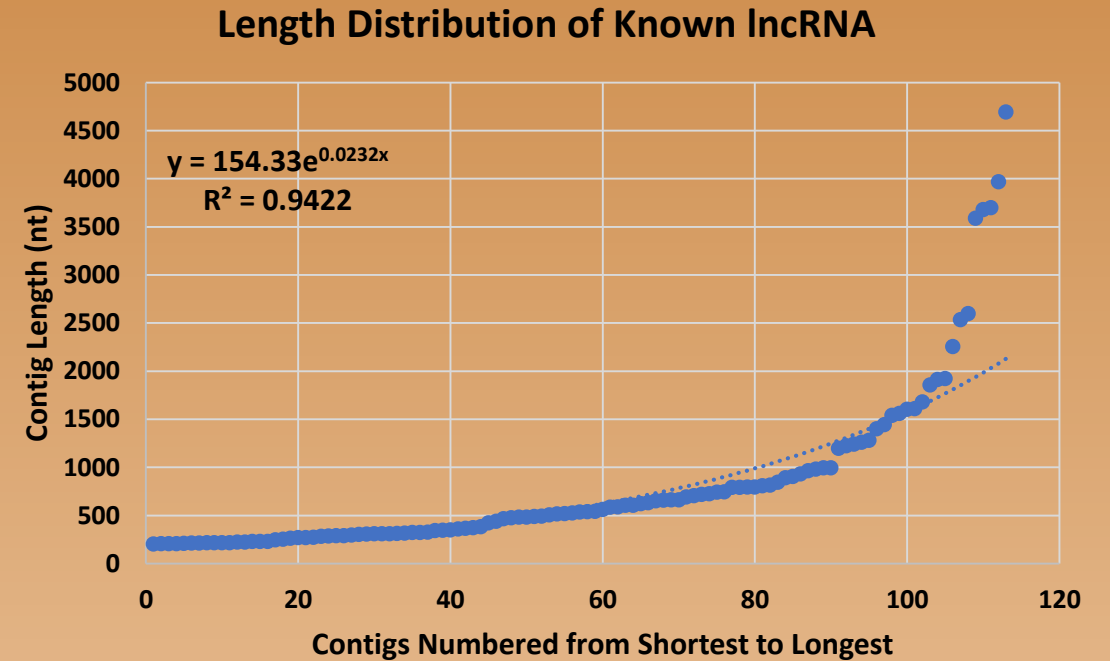
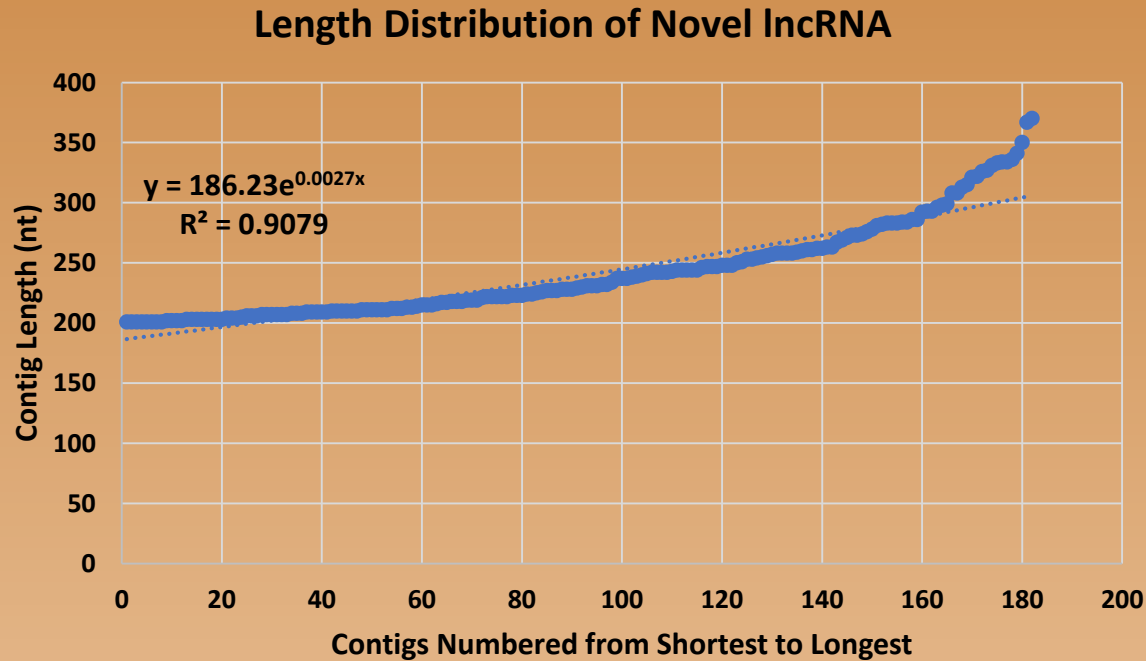
MFE vs Contig Length  
(Novel lncRNA)



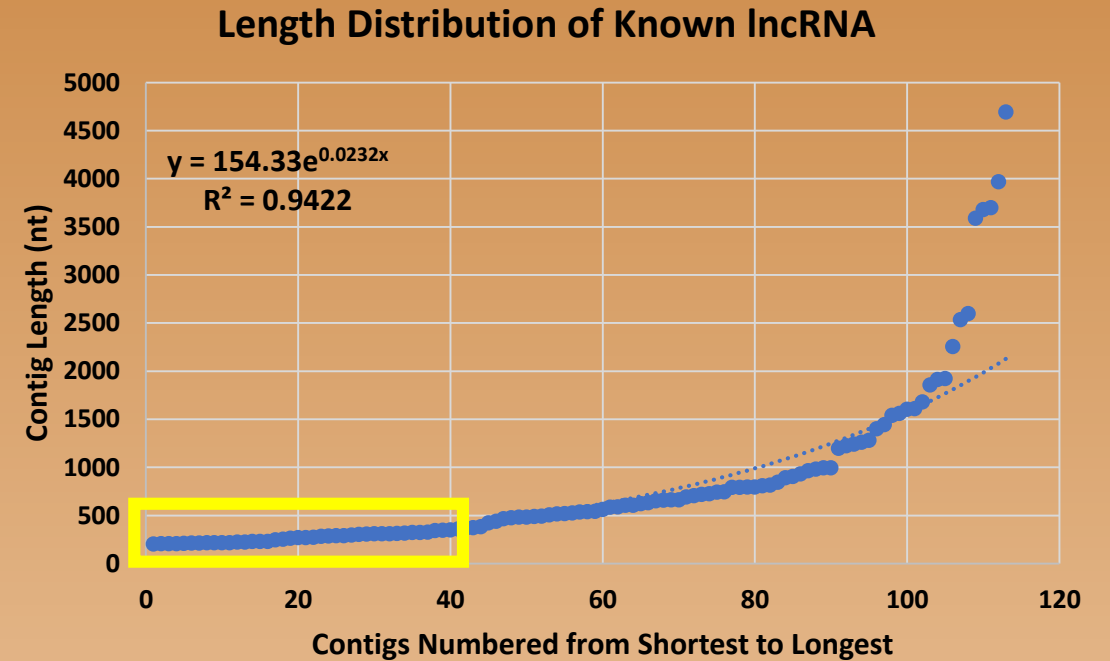
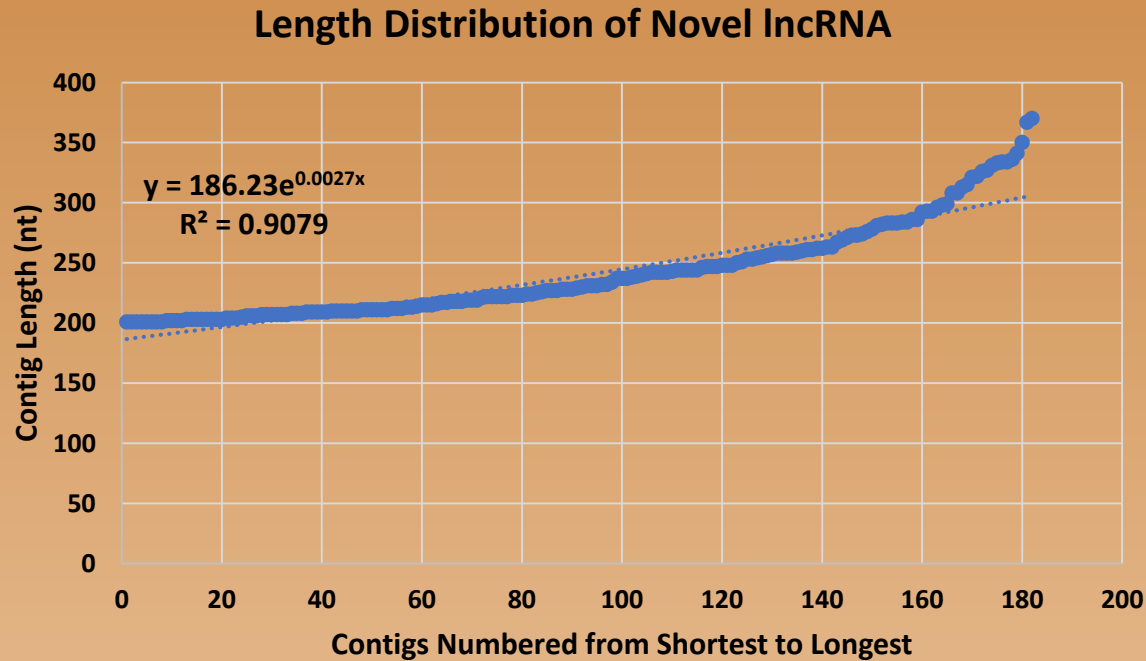
MFE vs Contig Length  
(Known lncRNA)



# The novel IncRNA represent the lower 40<sup>th</sup> percentile by length of the known IncRNA

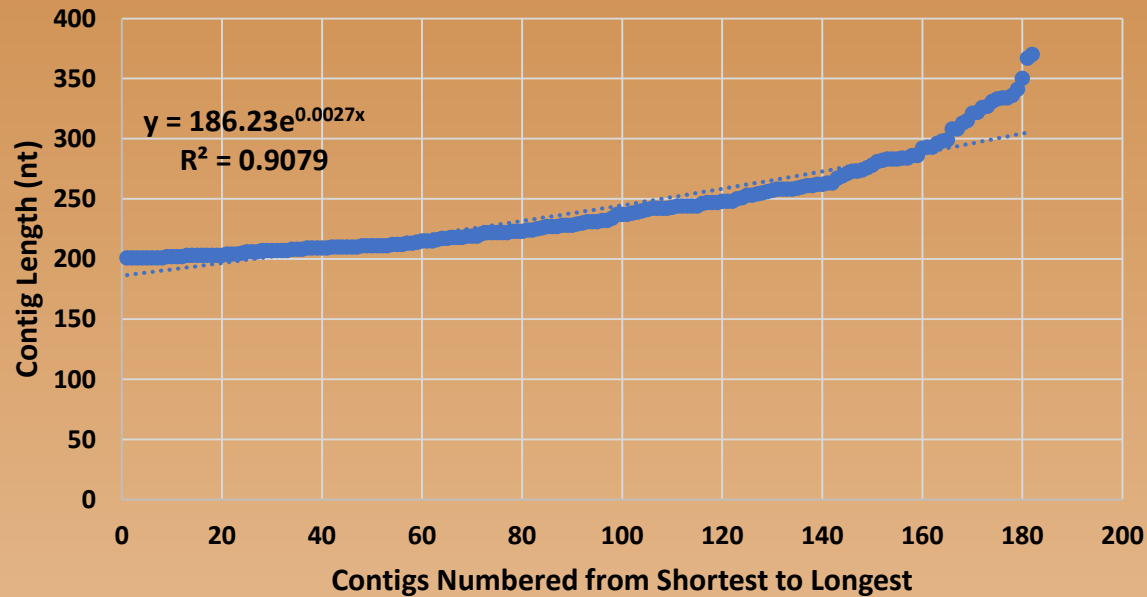


# The novel IncRNA represent the lower 40<sup>th</sup> percentile by length of the known IncRNA

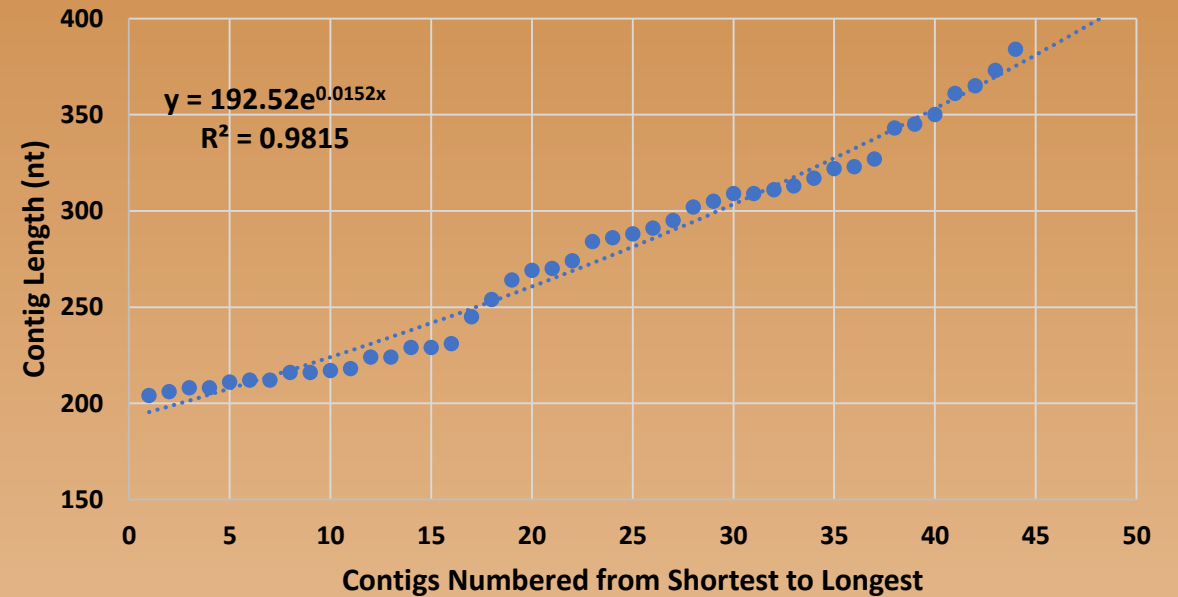


# The novel lncRNA have similar characteristics to the 40<sup>th</sup> percentile lncRNA

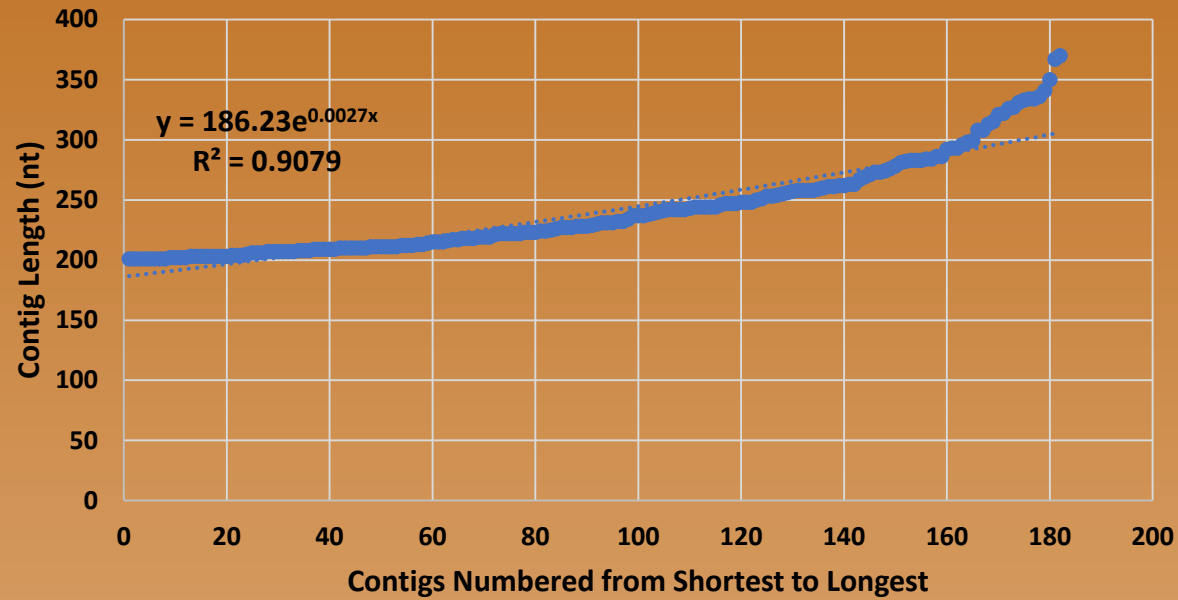
Length Distribution of Novel lncRNA



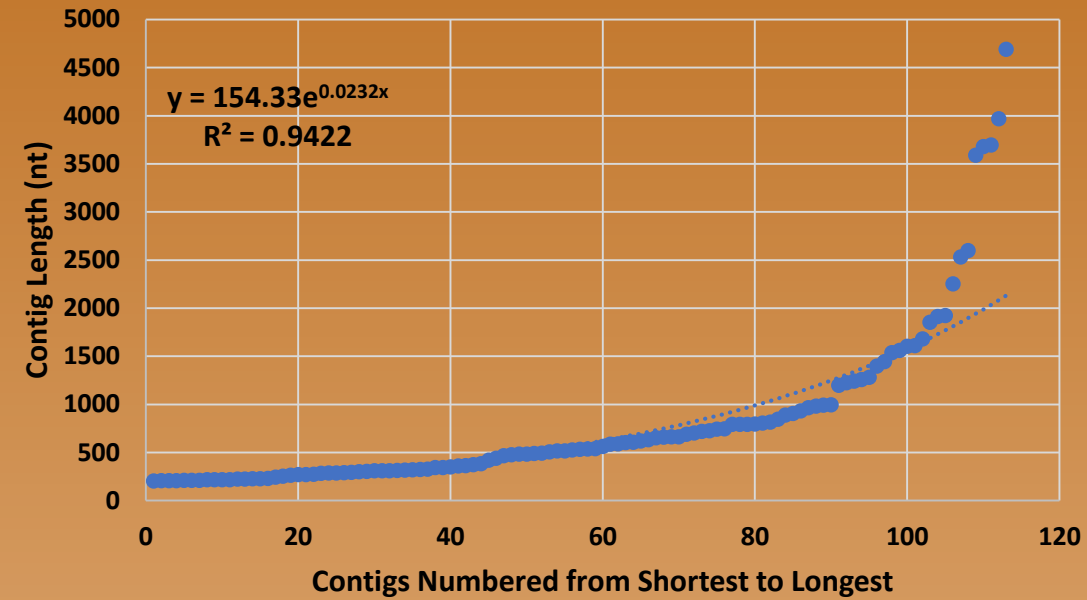
Length Distribution of 40th Percentile lncRNA



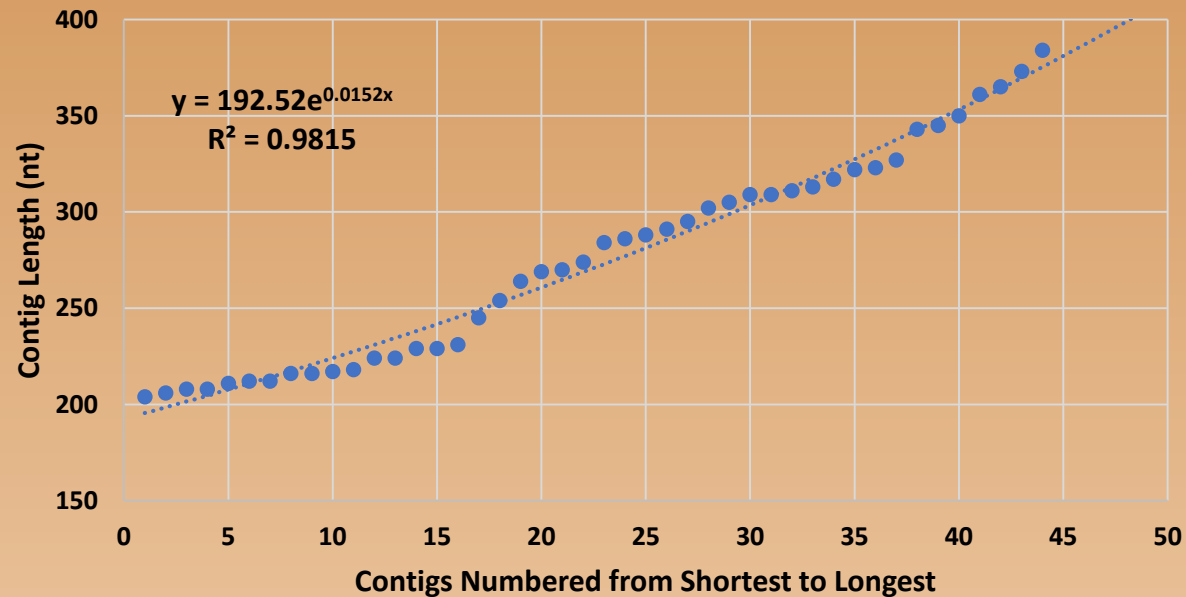
### Length Distribution of Novel IncRNA



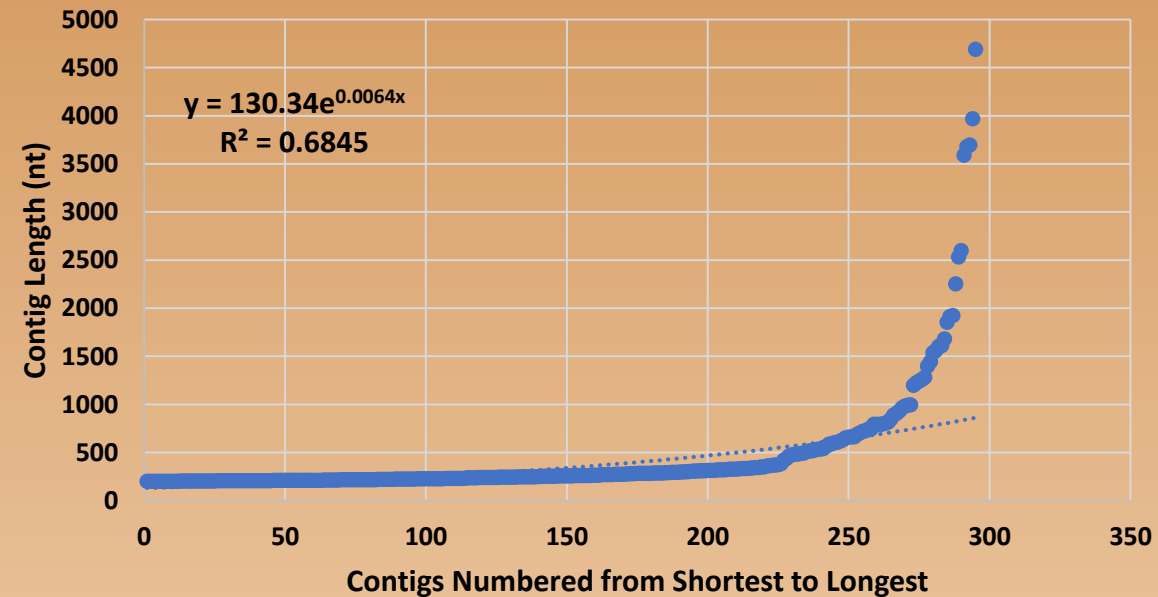
### Length Distribution of Known IncRNA



### Length Distribution of 40th Percentile IncRNA

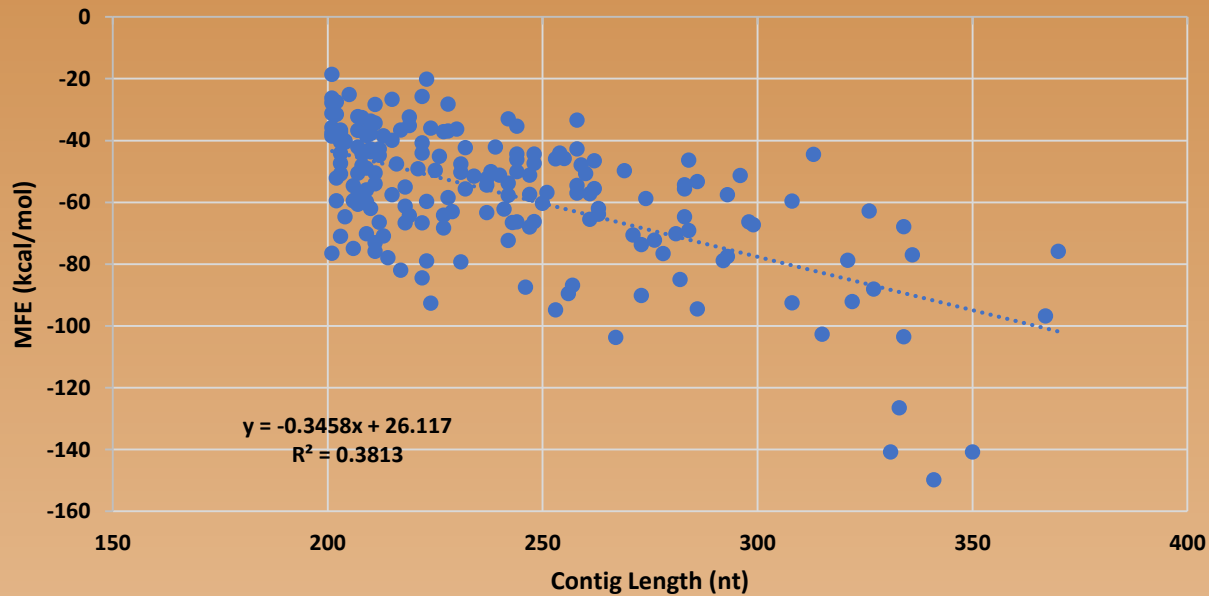


### Length Distribution of Pooled IncRNA

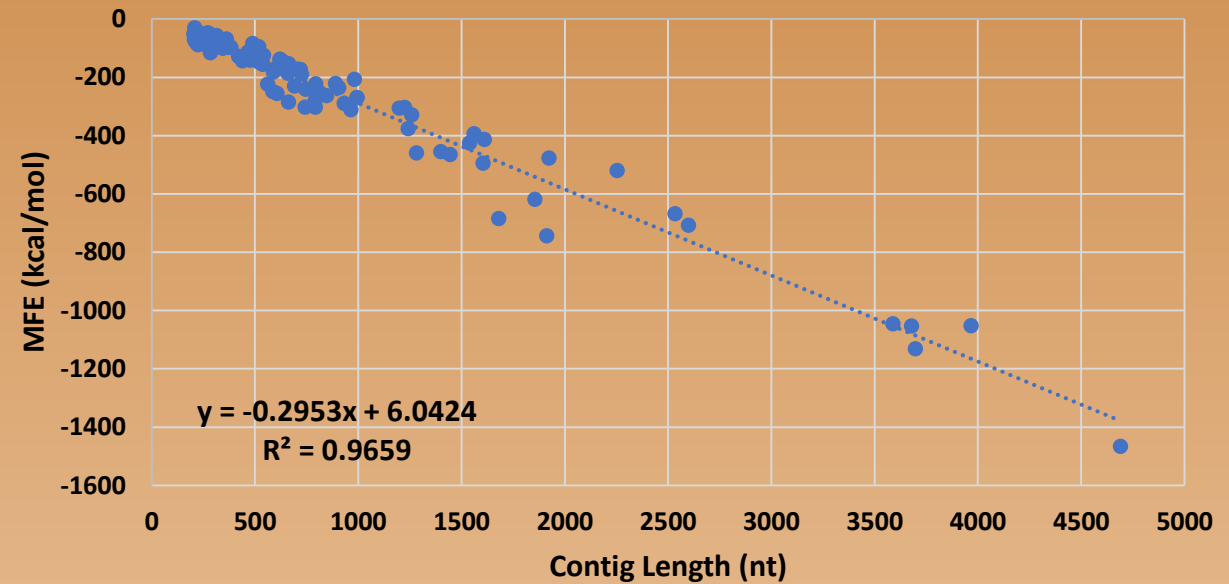


# The novel IncRNA have similar characteristics to the known IncRNA

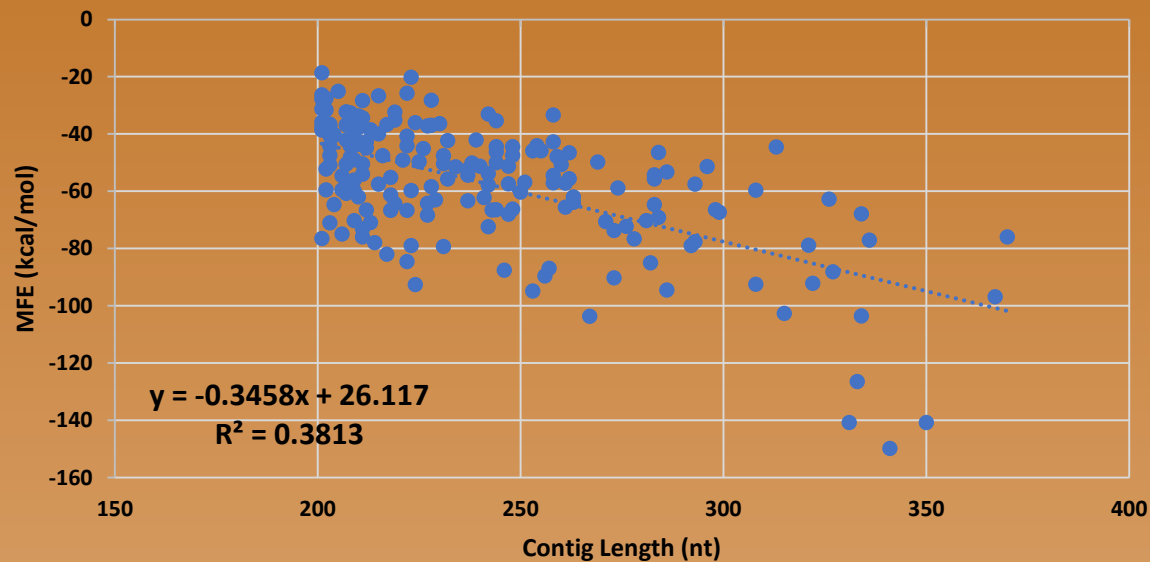
MFE vs Contig Length  
(Novel IncRNA)



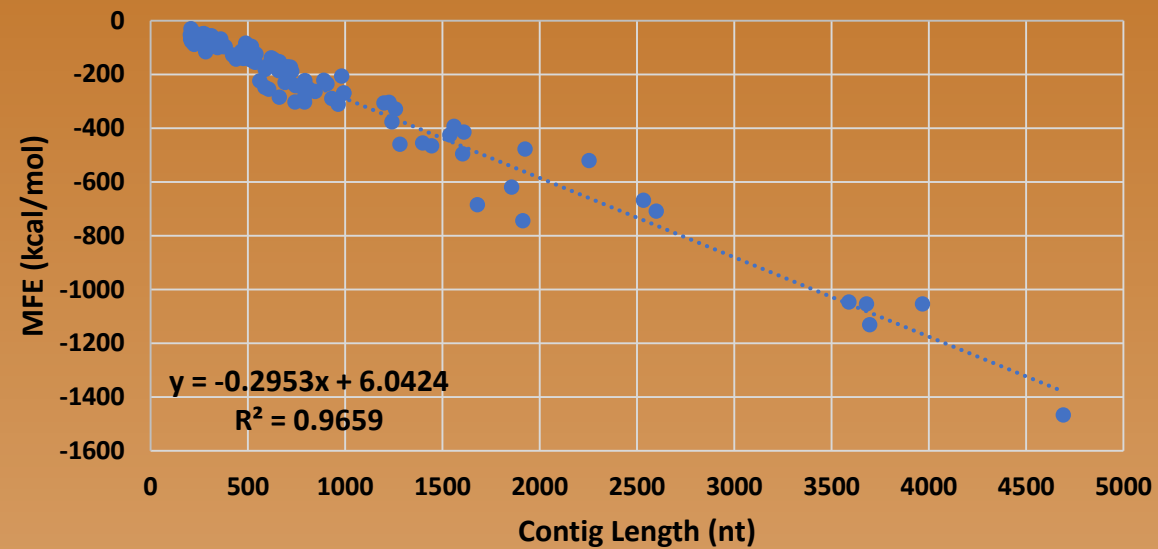
MFE vs Contig Length  
(Known IncRNA)



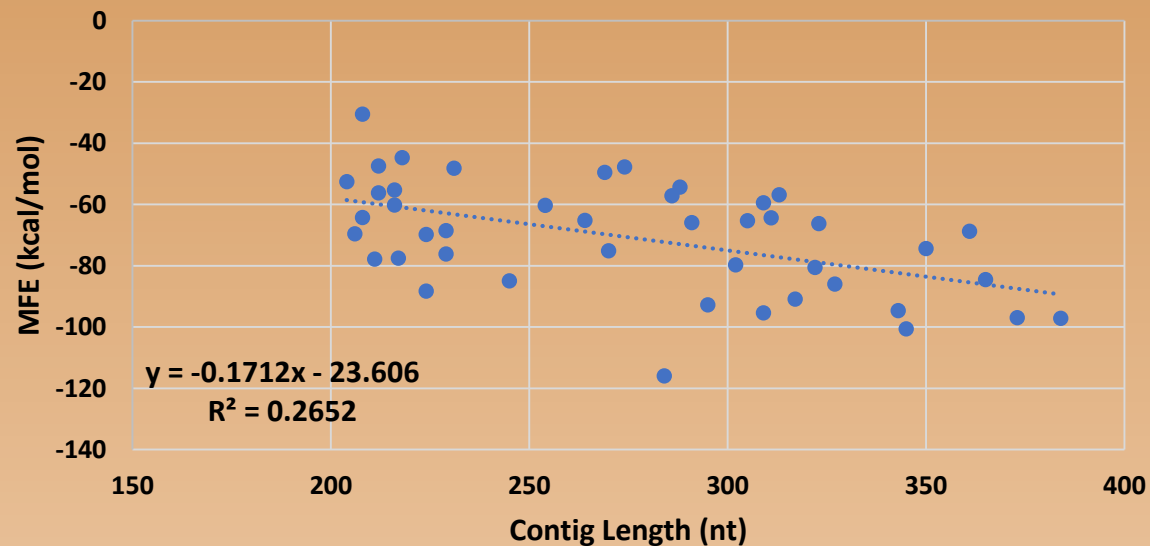
**MFE vs Contig Length  
(Novel lncRNA)**



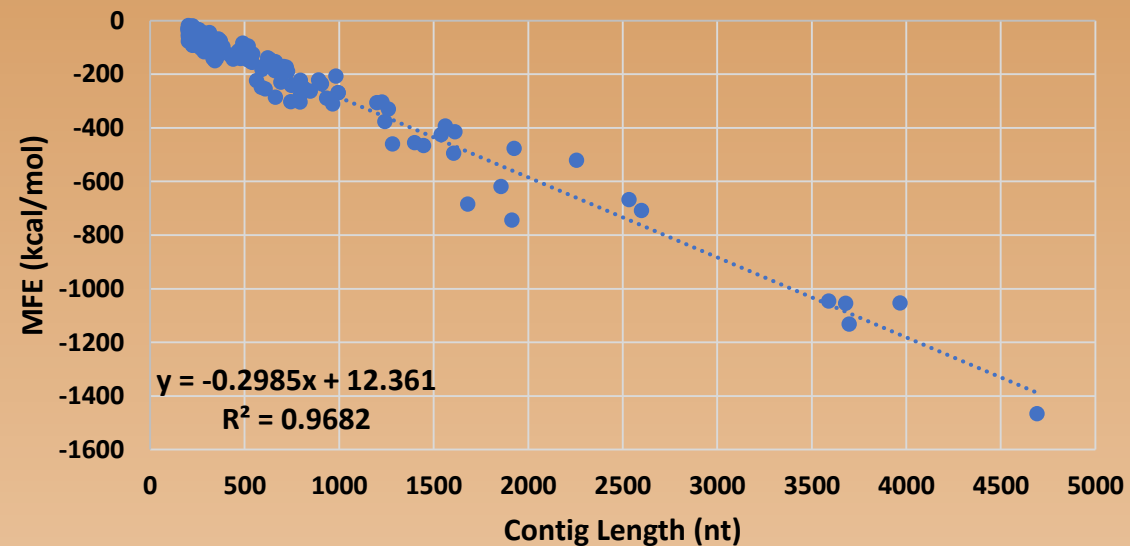
**MFE vs Contig Length  
(Known lncRNA)**



**MFE vs Contig Length  
(40th Percentile lncRNA)**



**MFE vs Contig Length  
(Pooled lncRNA)**



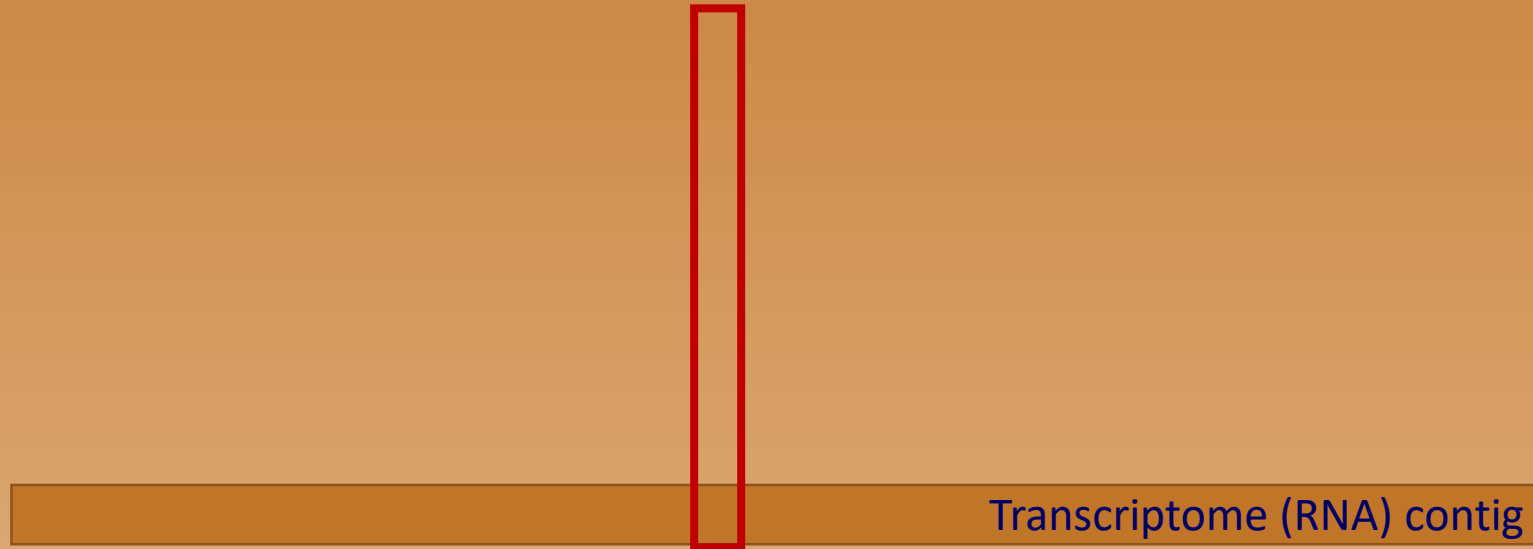




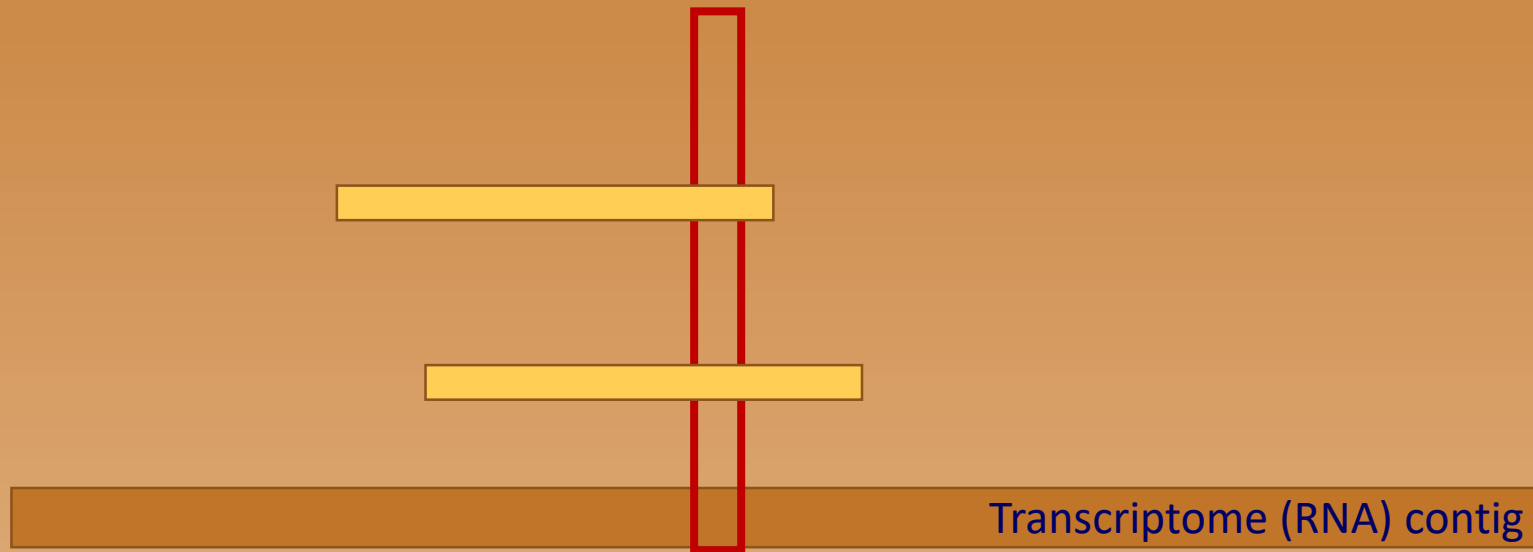
# We inspected coverage

Transcriptome (RNA) contig

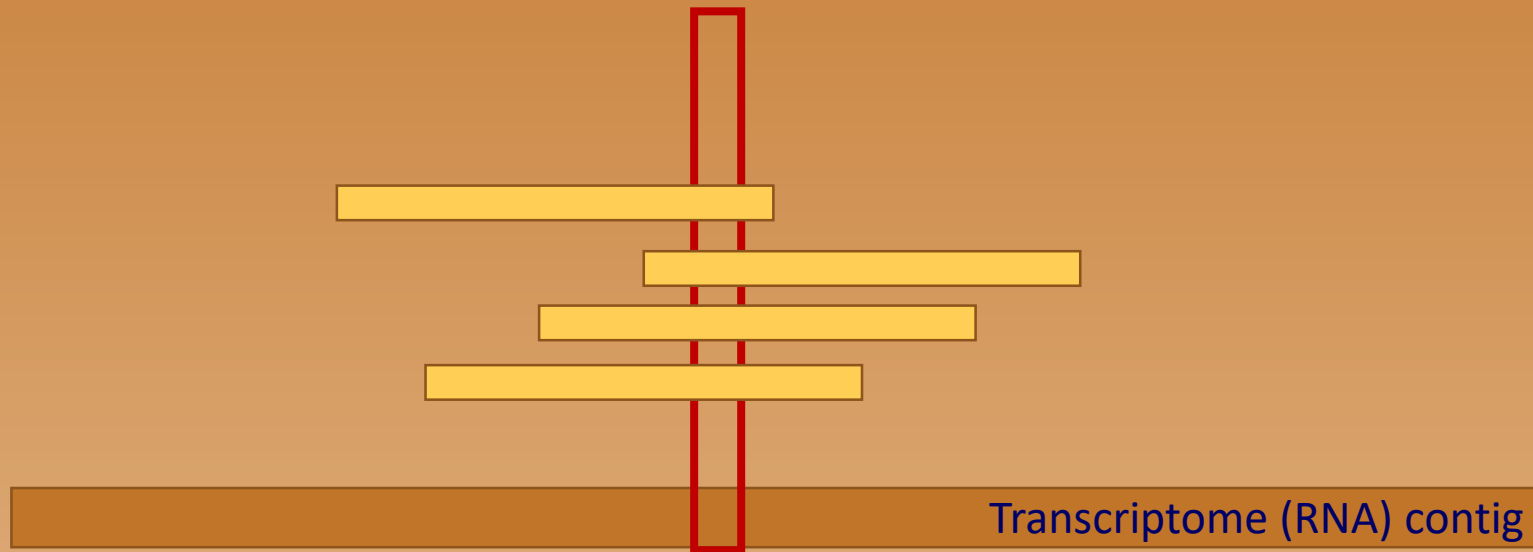
# We inspected coverage



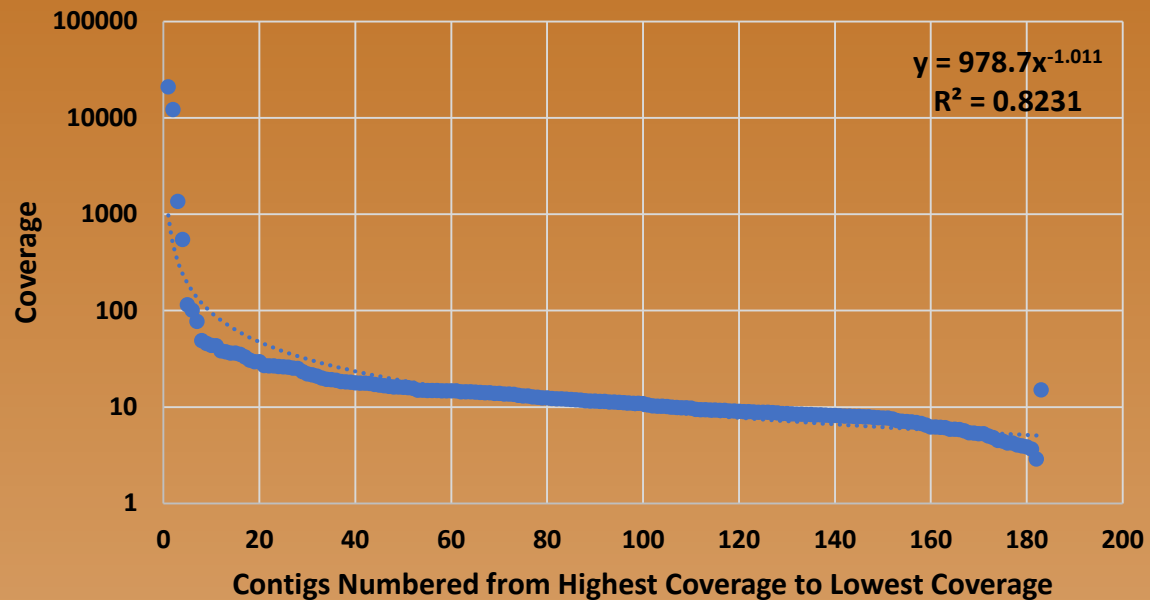
# We inspected coverage



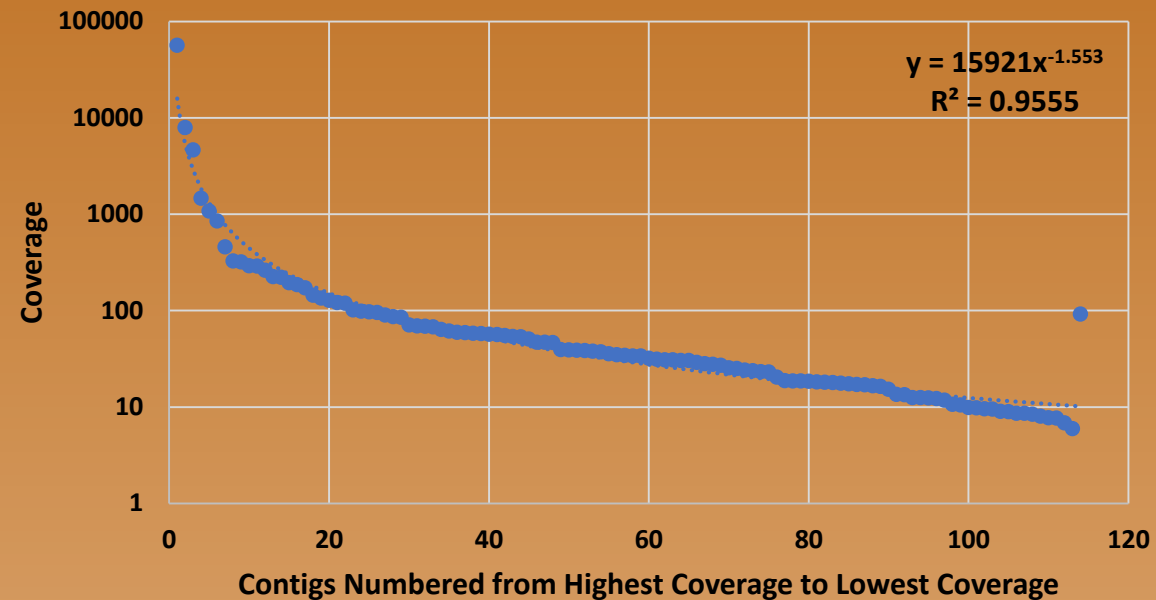
# We inspected coverage



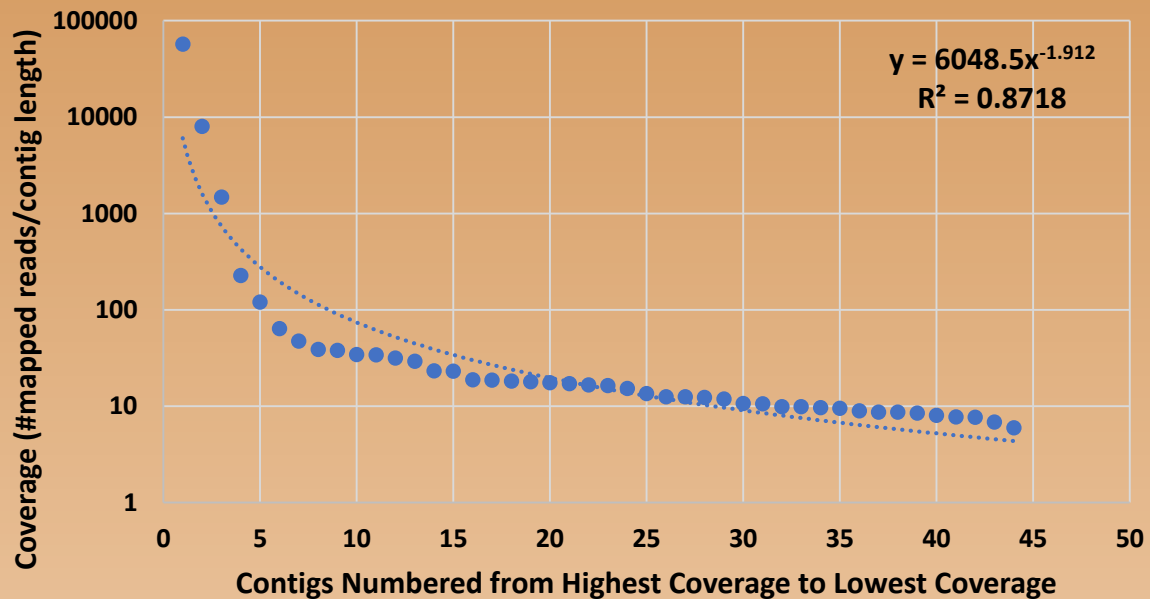
### Coverage Distribution of Novel IncRNA



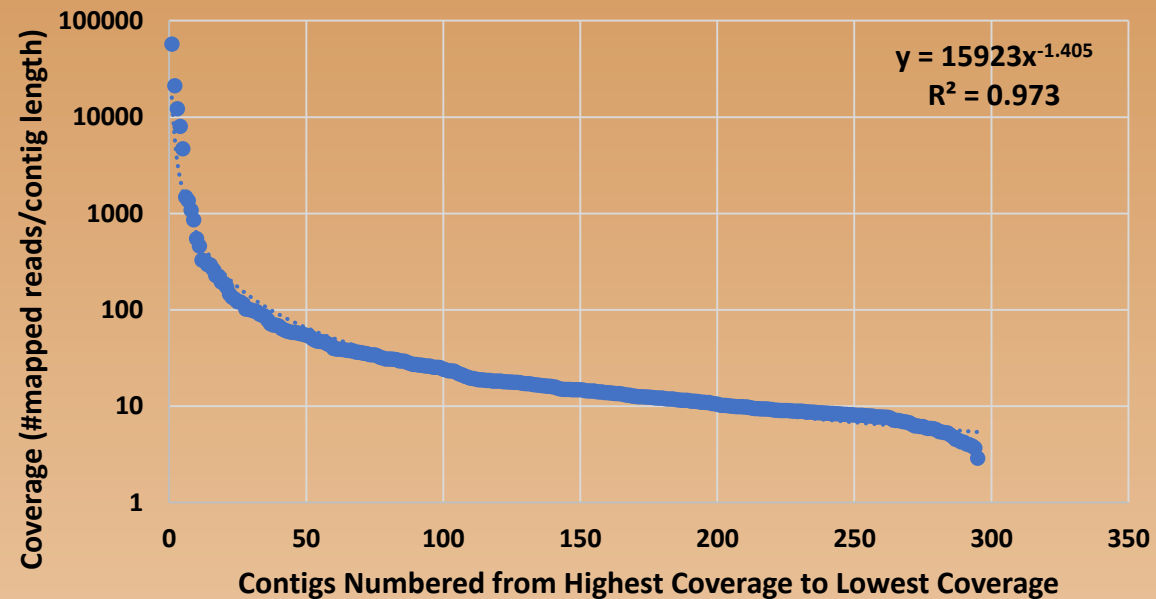
### Coverage Distribution of Known IncRNA



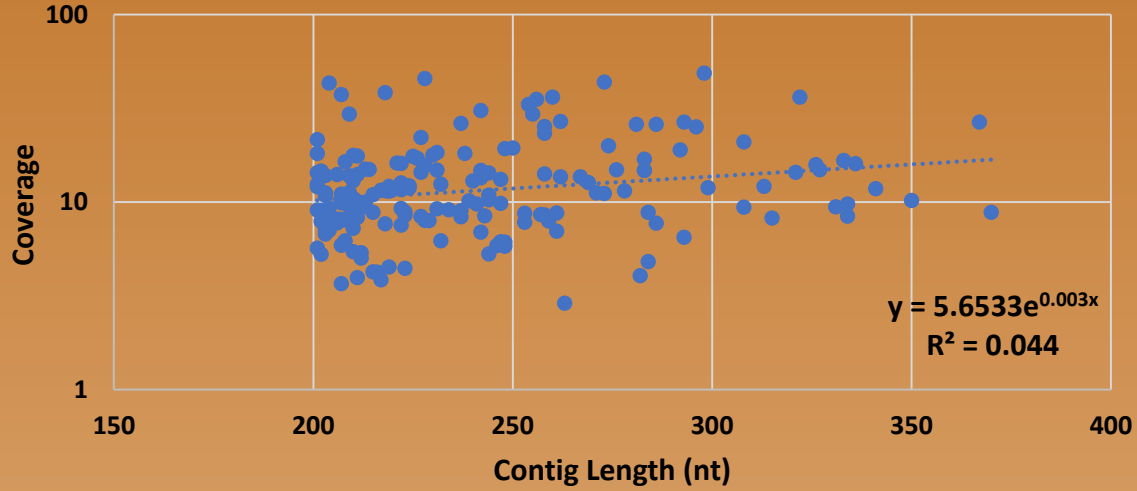
### Coverage Distribution of 40th Percentile IncRNA



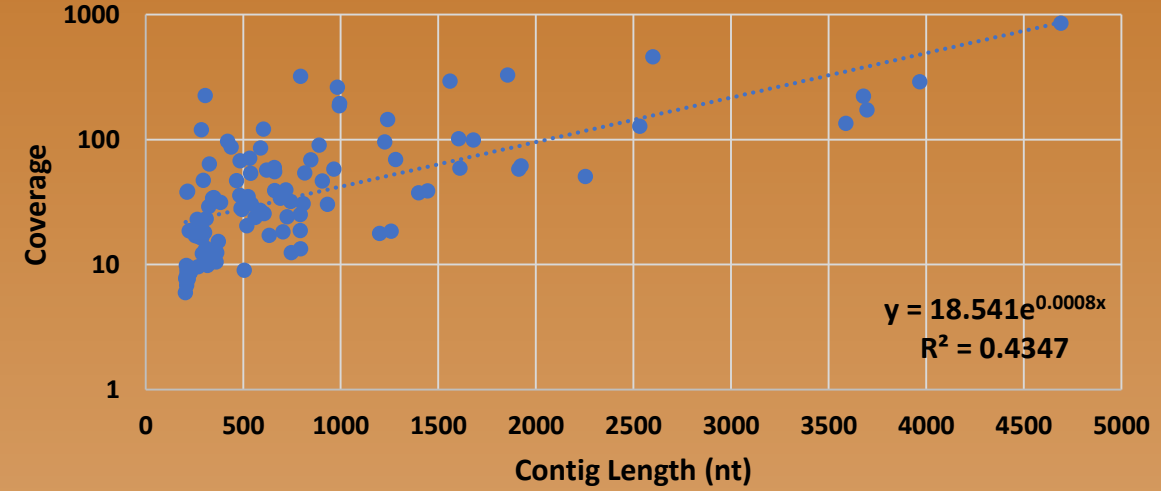
### Coverage Distribution of Pooled IncRNA



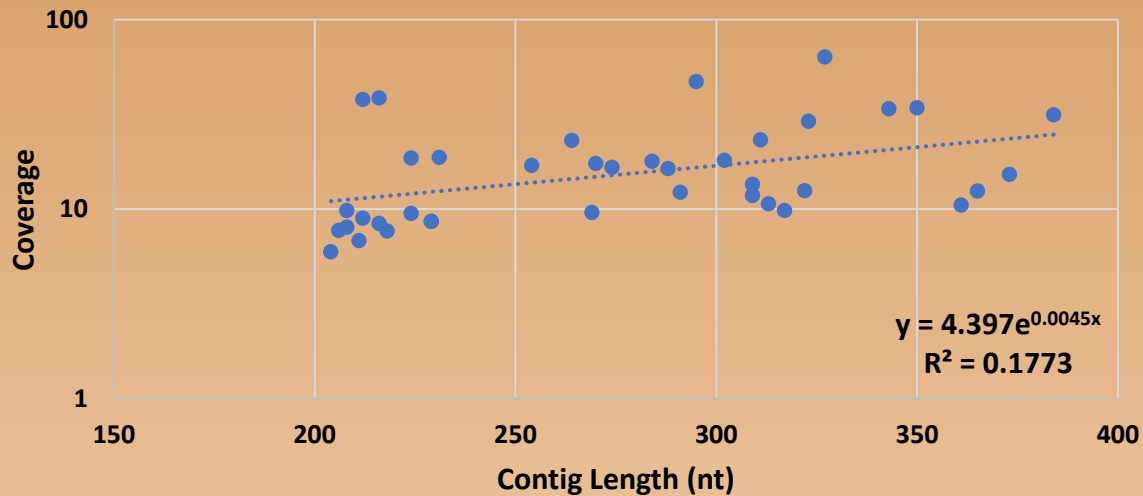
**Coverage vs Contig Length  
(Novel IncRNA)  
Without "Outliers"**



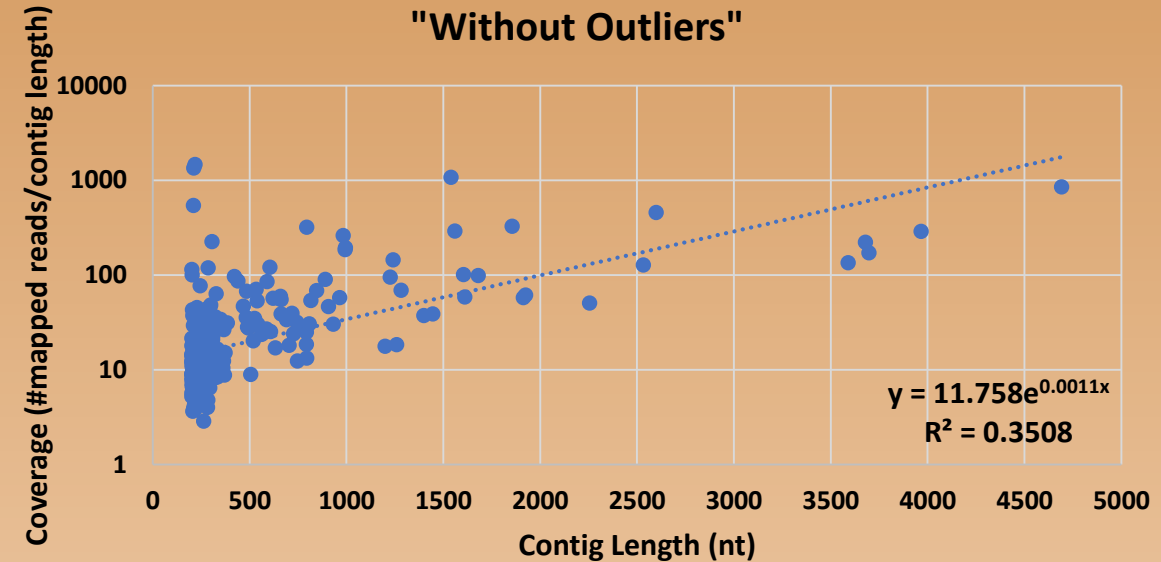
**Coverage vs Contig Length  
(Known IncRNA)  
Without "Outliers"**



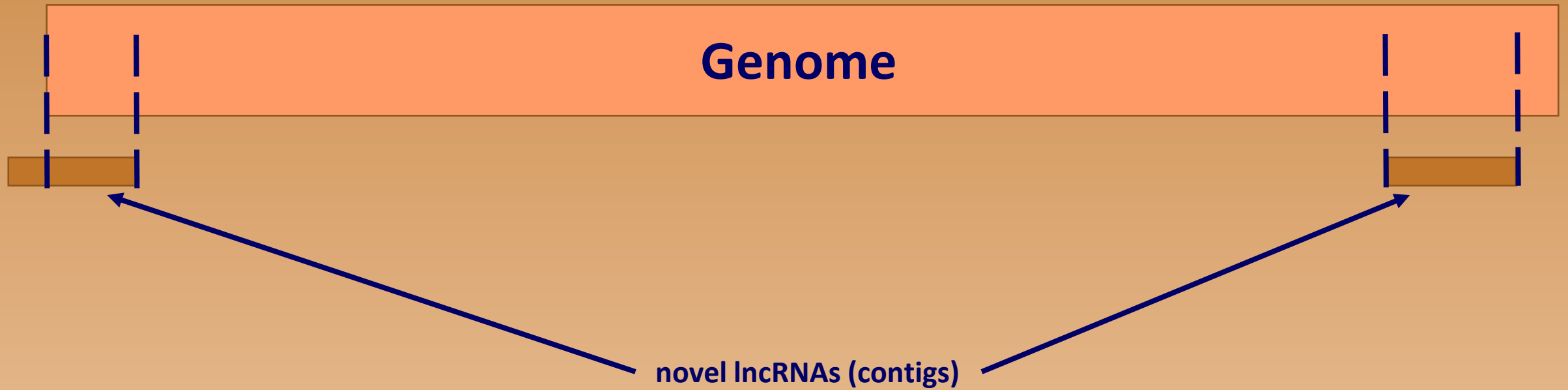
**Coverage vs Contig Length  
(40th Percentile IncRNA)  
Without "Outliers"**



**Coverage vs Contig Length  
(Pooled IncRNA)  
"Without Outliers"**

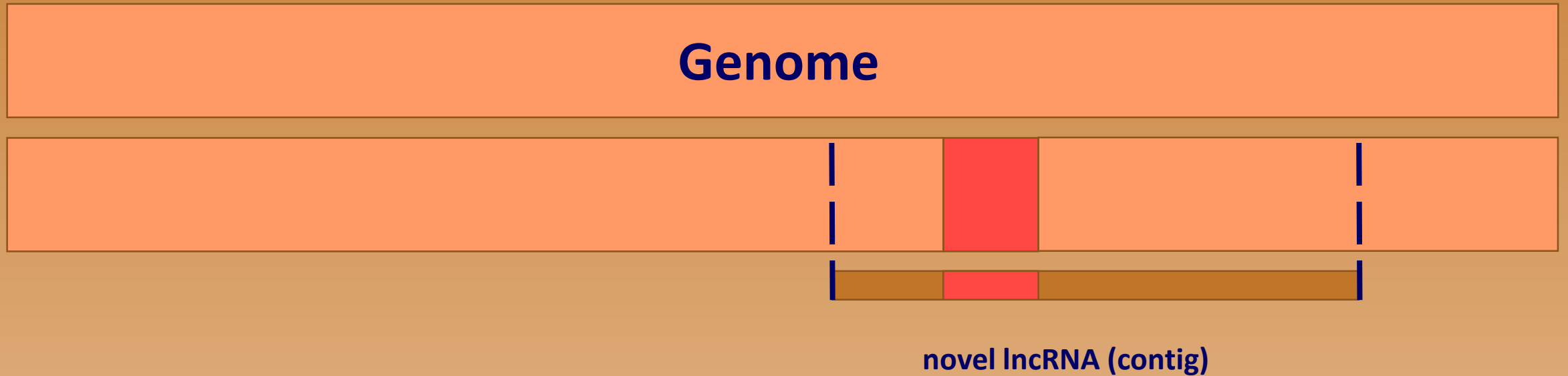


# The novel lncRNA map to the genome

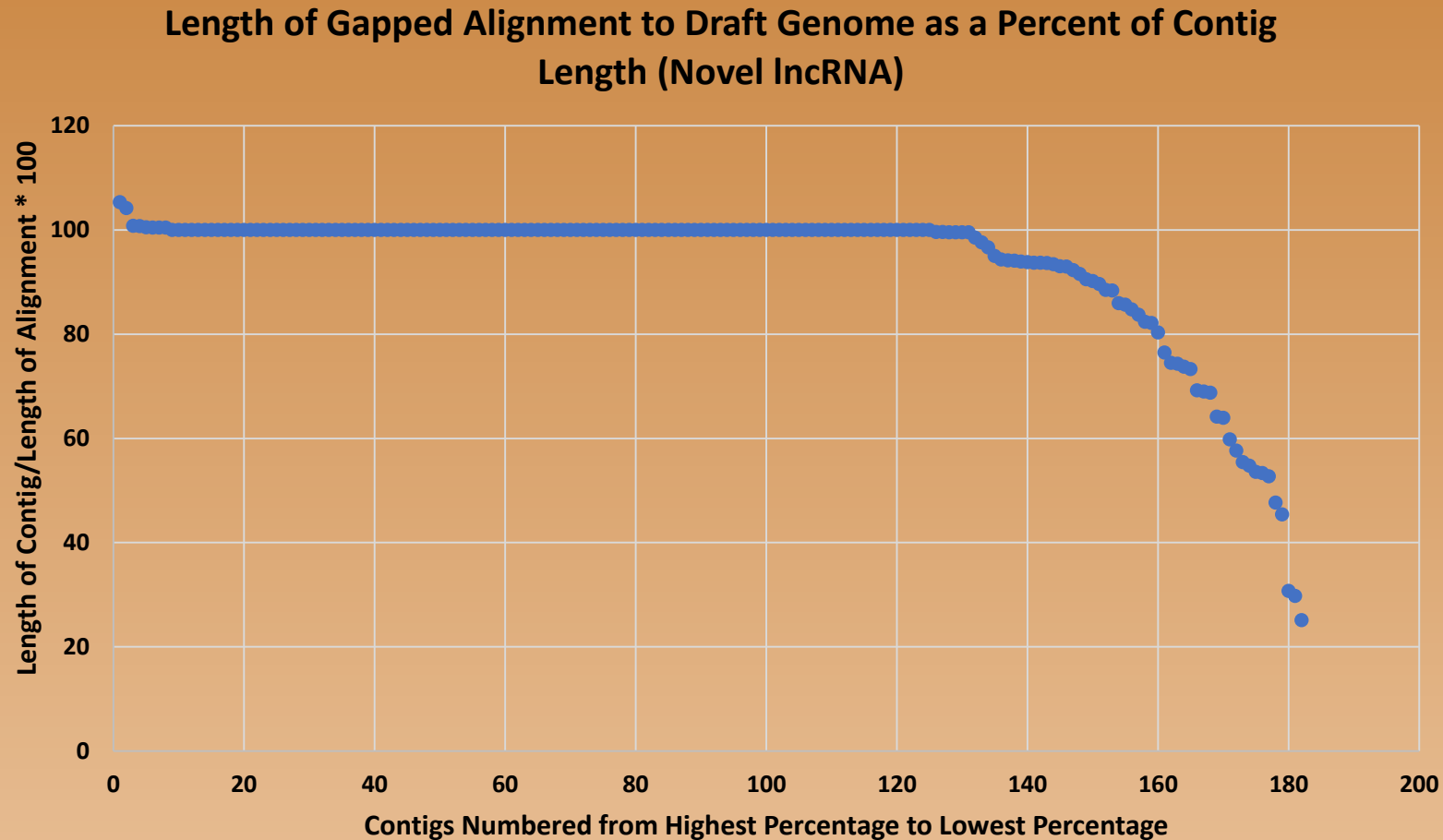




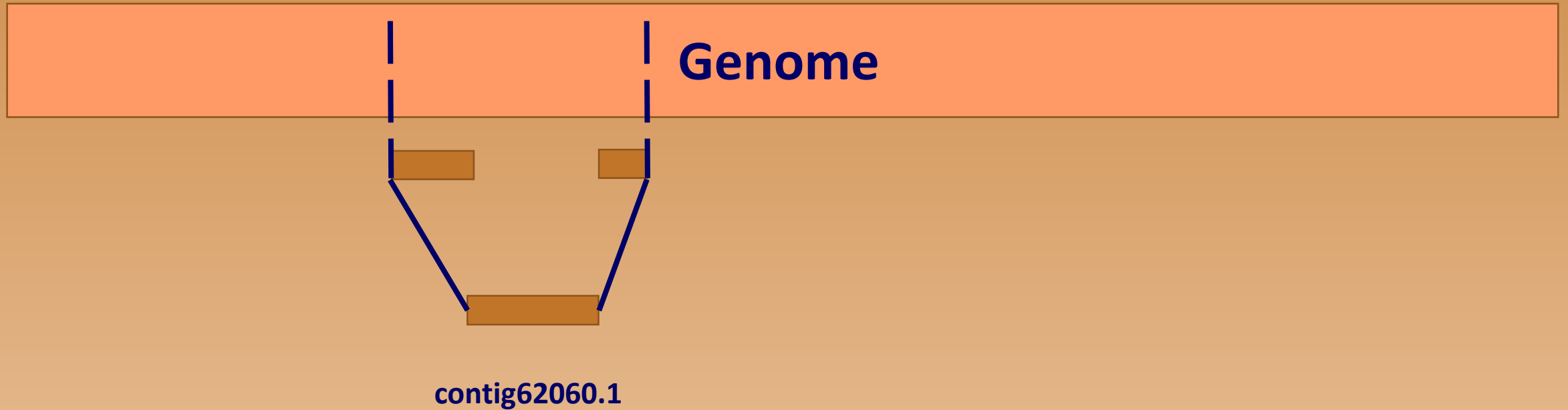
# The match is not necessarily perfect



# Most contigs aligned with the genome along most of the contig's length



# One novel lncRNA is potentially multi-exonic



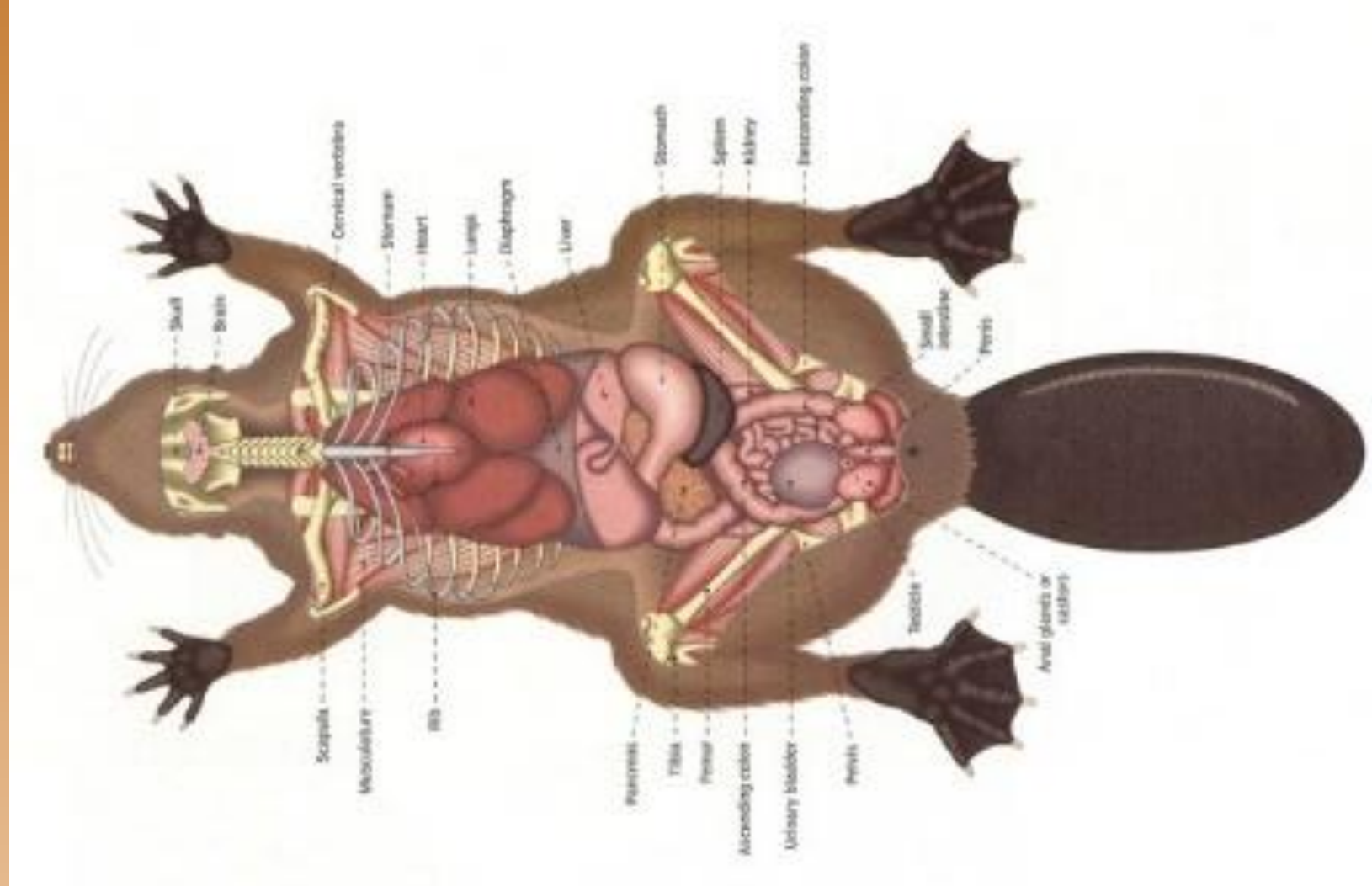
# Nine lncRNAs particularly warrant further investigation

Contig	Measure				
	Length	MFE	Coverage	BLASTn alignment length	Intronic
contig41254.1					
contig43610.1					
contig44966.1					
contig46102.1					
contig45799.1					
contig46542.1					
contig46174.1					
contig59927.1					
contig62060.1					

# Overview

- Background
- Objective and Impact
- Methodology
- Results
- **Further Research**

# The tissue atlas will help find tissue specific lncRNA



# Acknowledgments

## Oregon Zoo

Mitch Finnegan

Tim Storms

Amy Cutting

David Shepherdson



THANK YOU