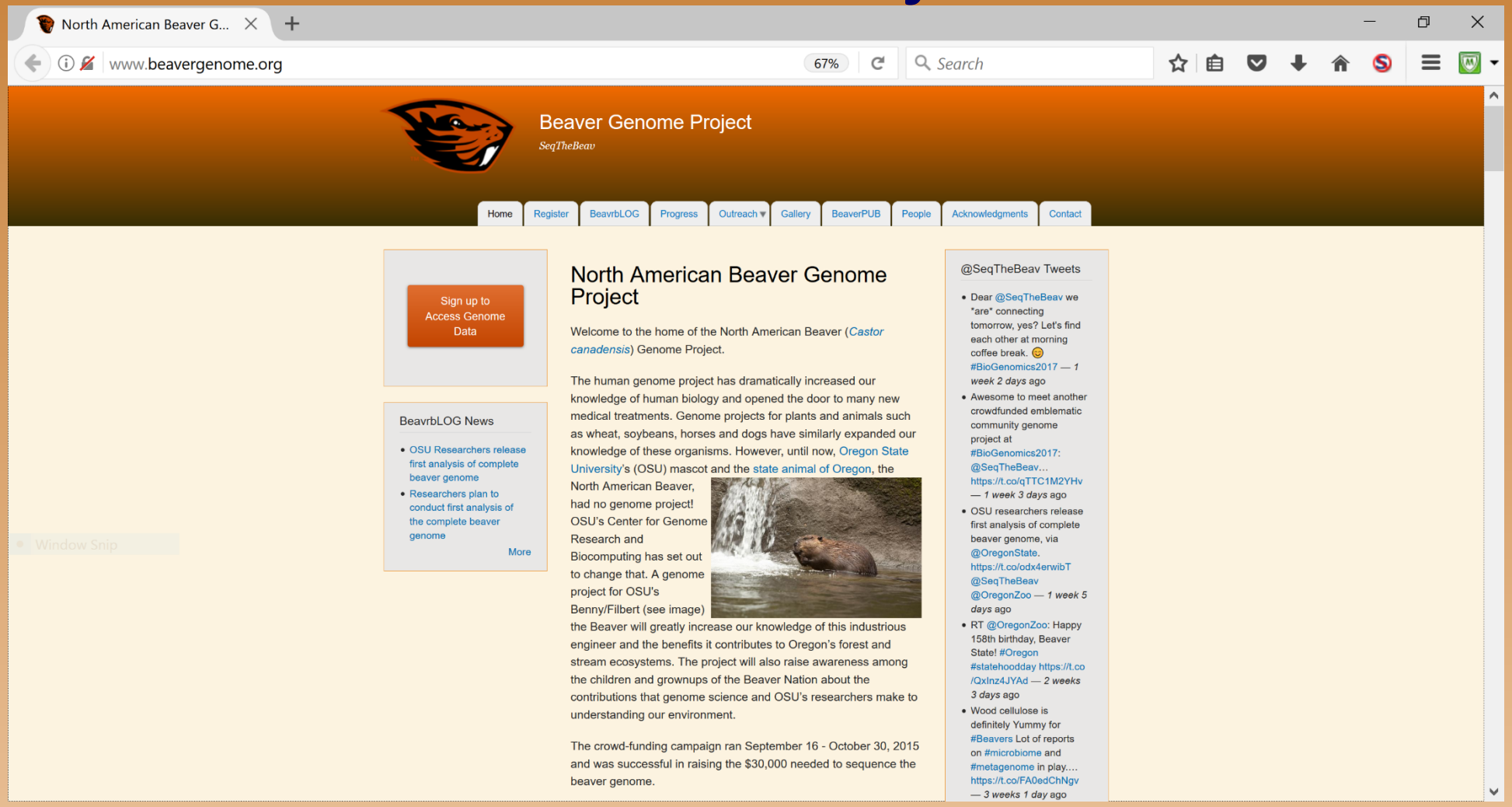
Long Noncoding RNA in the North American Beaver

Amita Kashyap[1, 4]; Adelaide C Rhodes[2, 4]; and Stephen A Ramsey, PhD[3, 4]

[1]BioResource Research, College of Agricultural Sciences
[2]Center for Genome Research and Biocomputing
[3]Department of Biomedical Sciences
[4]Oregon State University, Corvallis

Image Credit: Dreams of Animals. animal-dream.com/beaver.html (2015)

This research is part of OSU's Beaver Genome Project



Overview

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

Overview

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

The beaver has many unique characteristics

Image Credit: Préau, L.-M. *Beavering*. (2011).

The Central Dogma is not the whole story

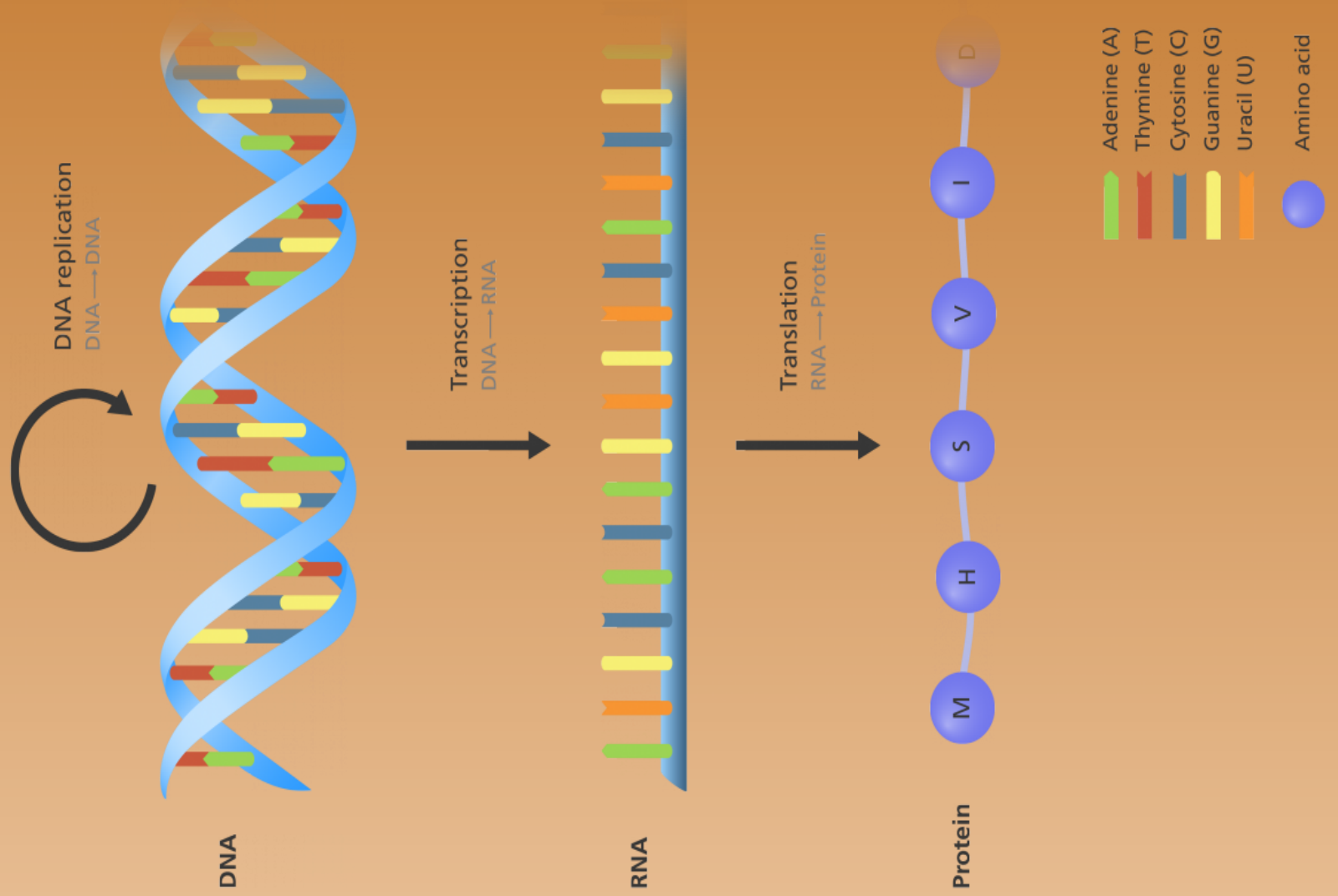


Image Credit: Genome Research Limited. *An illustration showing the flow of information between DNA, RNA and protein.* (2016).

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

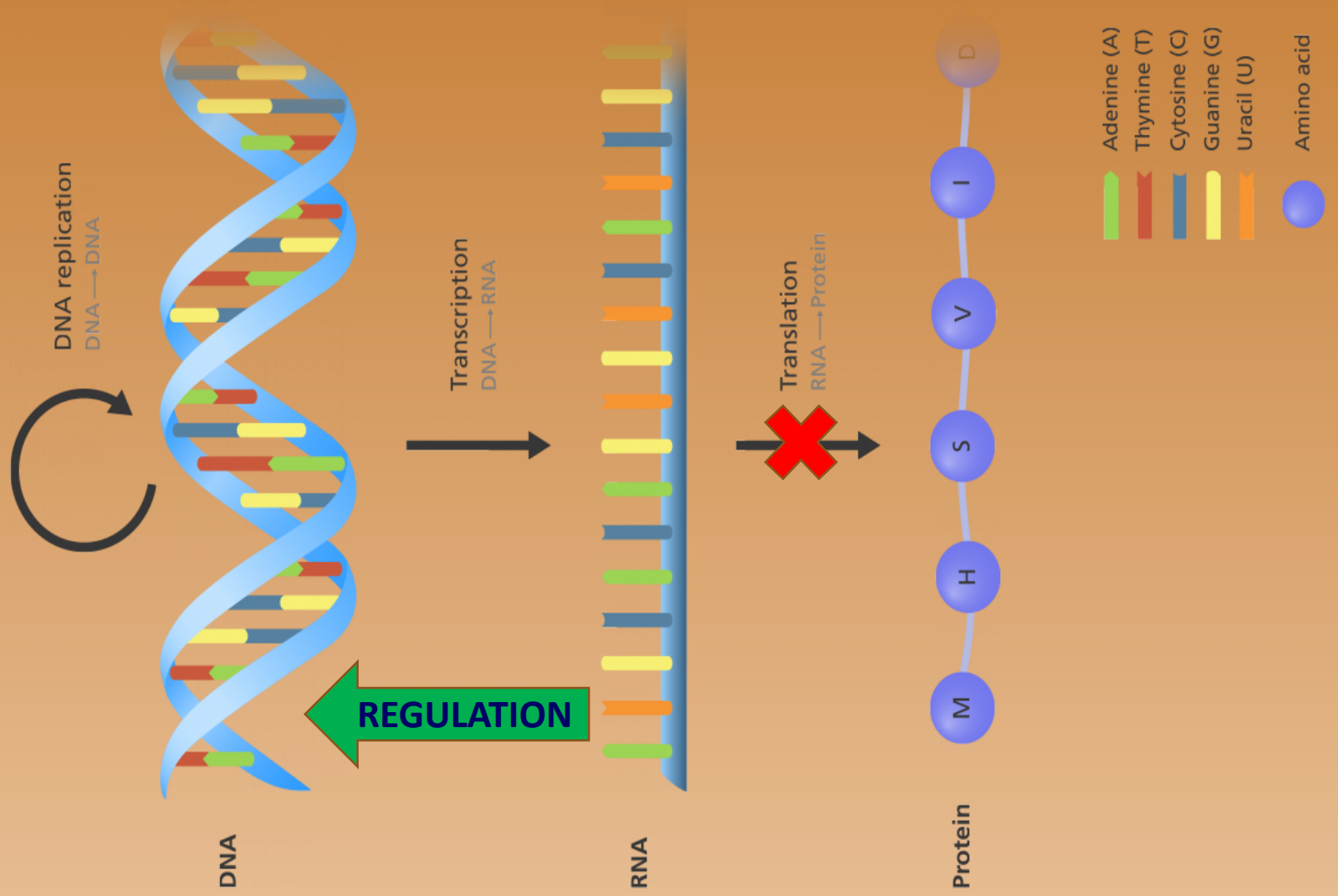
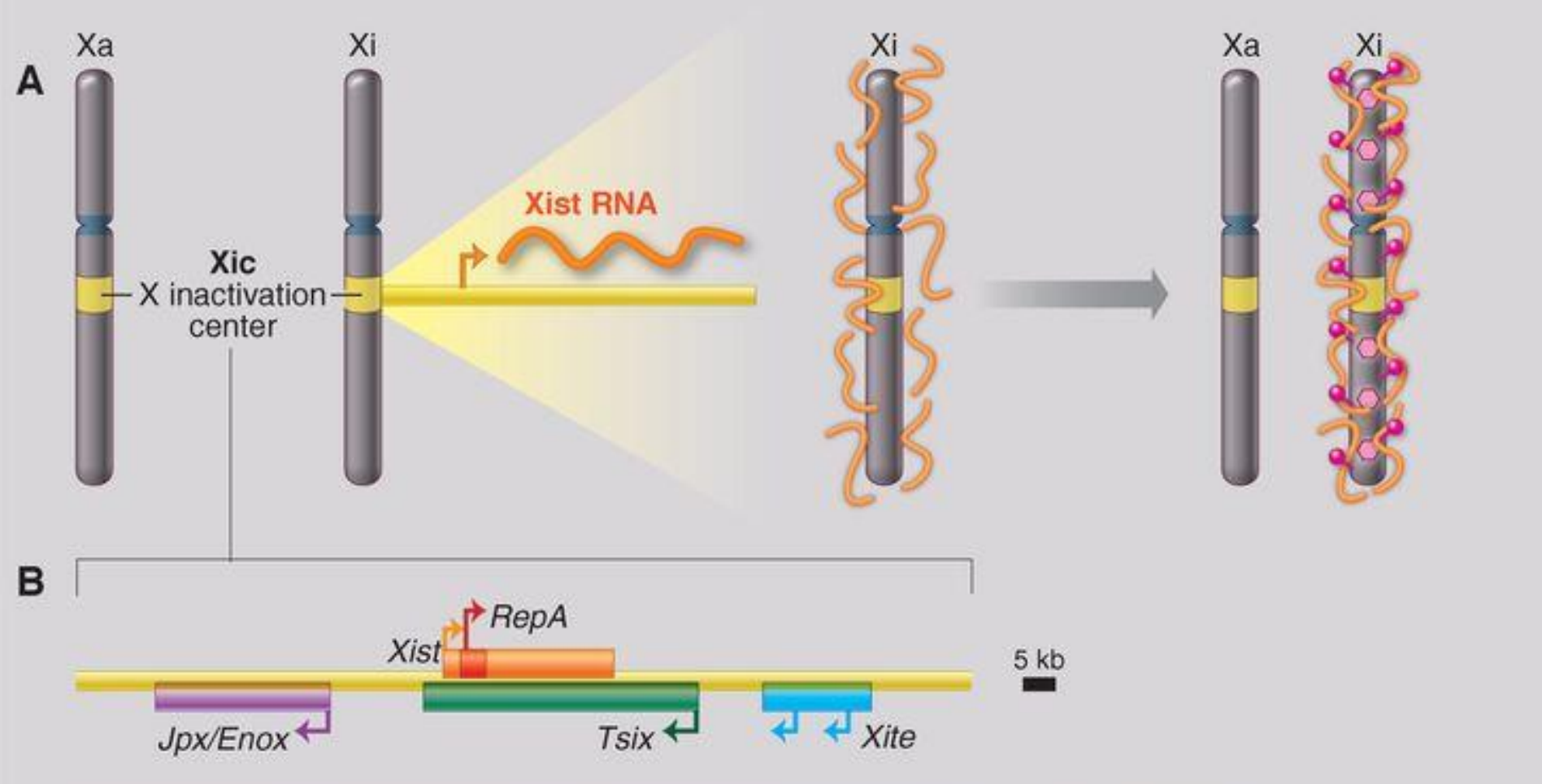


Image Credit: Genome Research Limited. *An illustration showing the flow of information between DNA, RNA and protein.* (2016).

Long noncoding RNA (lncRNA) is regulatory



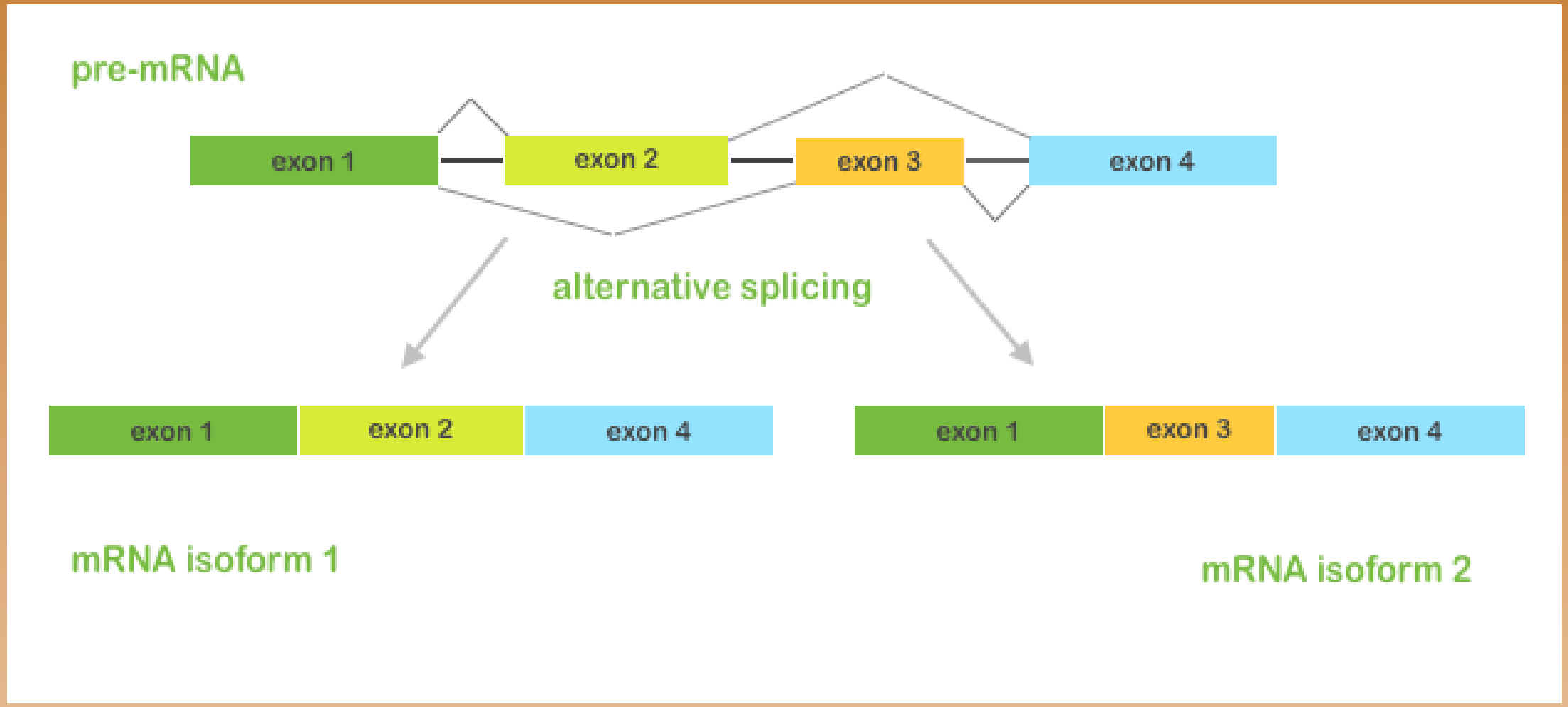
Lee, J. T. Epigenetic Regulation by Long Noncoding RNAs. *Science* **338,** 1435–1439 (2012).

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)

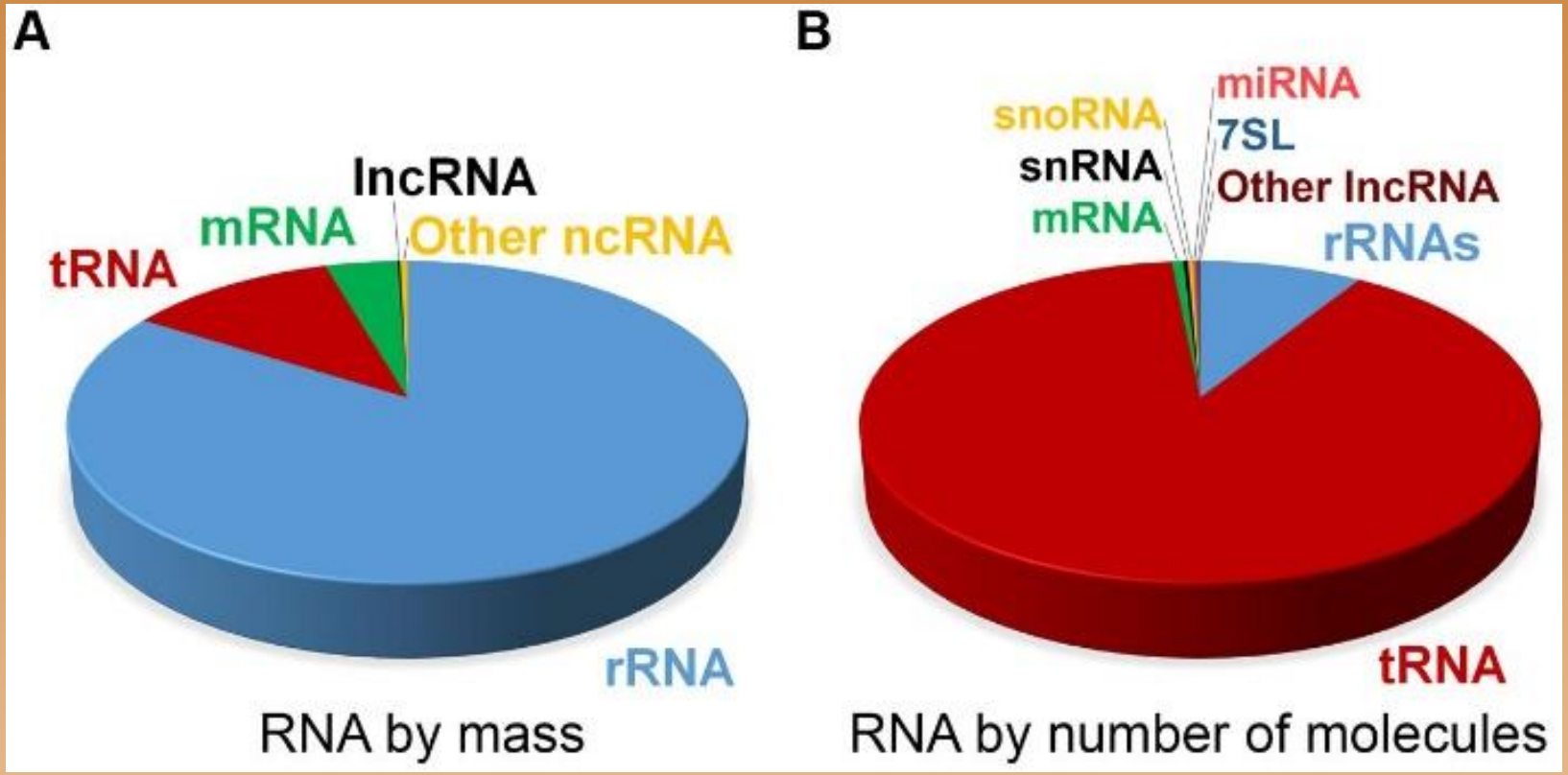
Amaral, P. P., Dinger, M. E. & Mattick, J. S. Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective. *Briefings in Functional Genomics* **12,** 254–278 (2013).

The transcriptome is the sum total of RNA



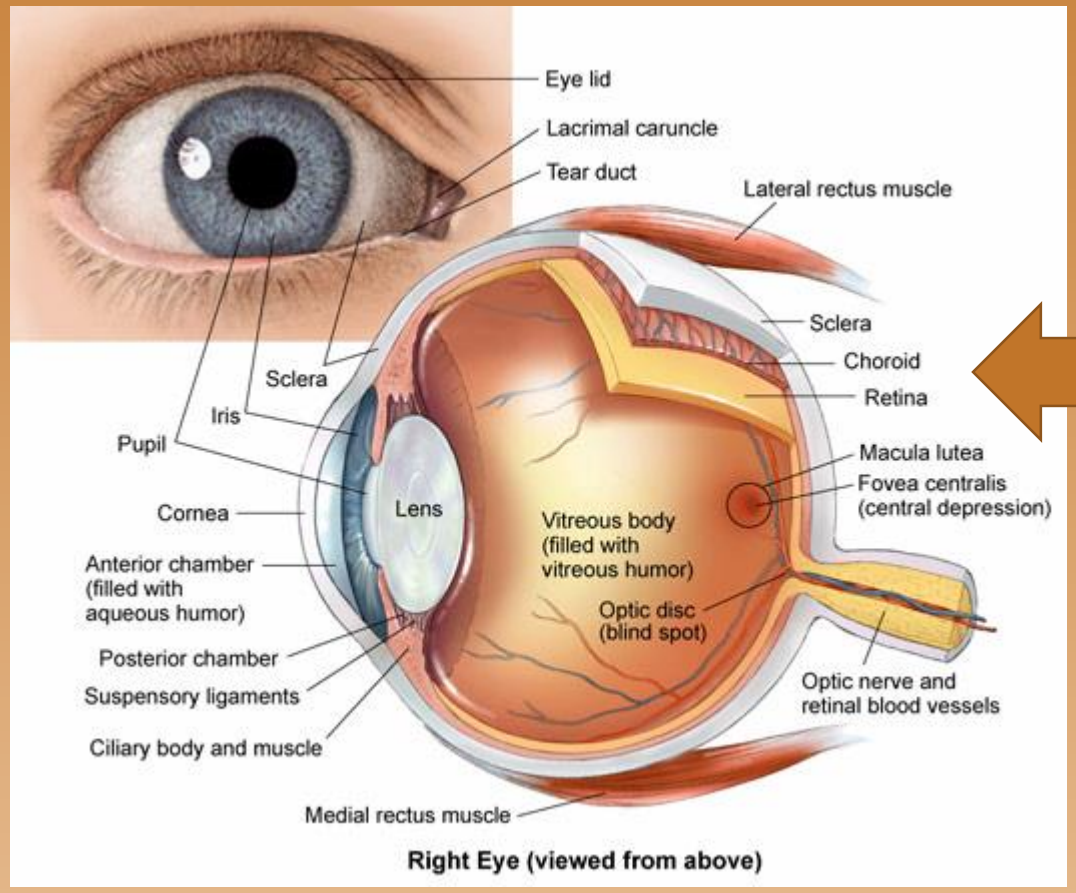
Golovina, E. Testing Differential Isiform Expression on Genestack Platform. *Genestack* (2015).

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



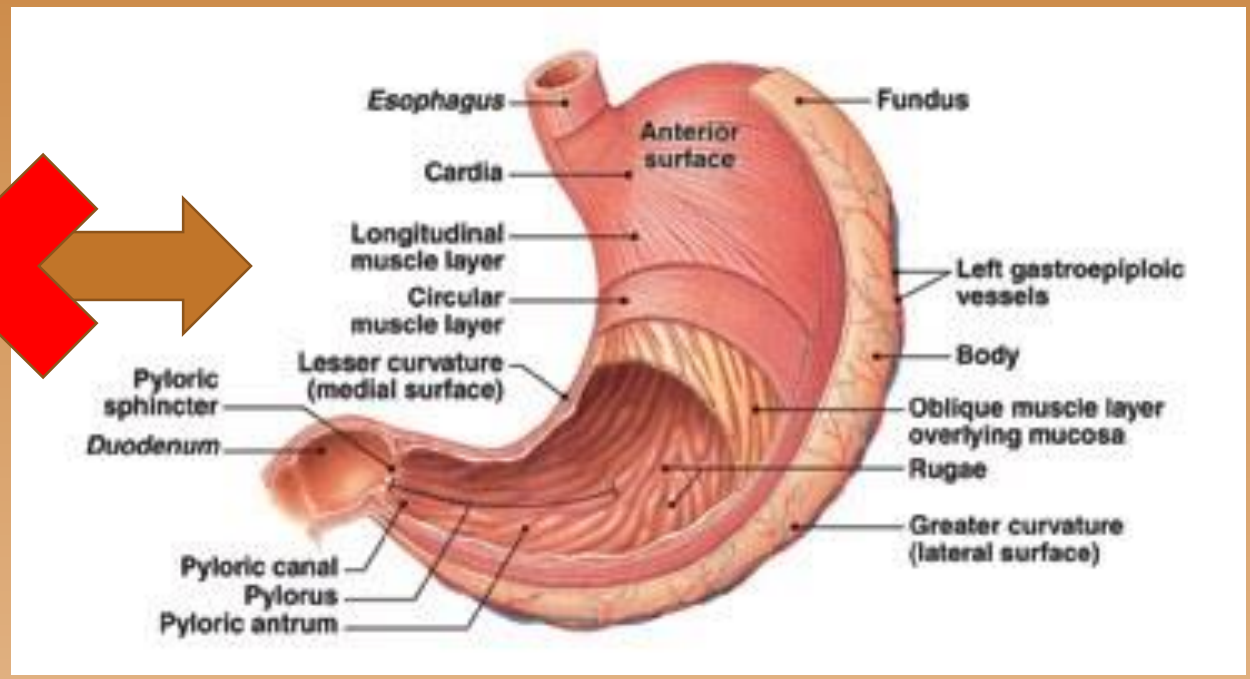


Image Credit: OrgansOfTheBody.com. *Stomach Anatomy*
Image Credit: Your Eyes Center. *Eye Anatomy*

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

Image by Michael Durham, courtesy Oregon Zoo

RNA came from 16 tissues

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

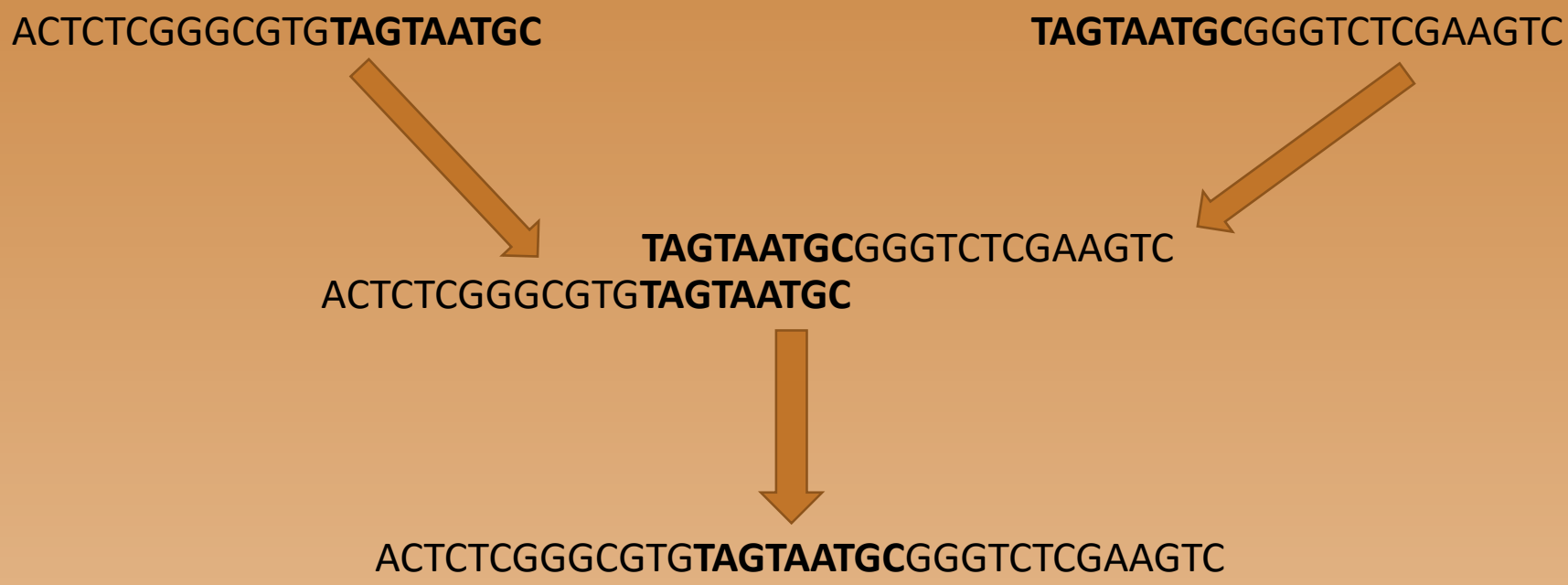
Assemblers assemble by overlap

ACTCTCGGGCGTGTAGTAATGC

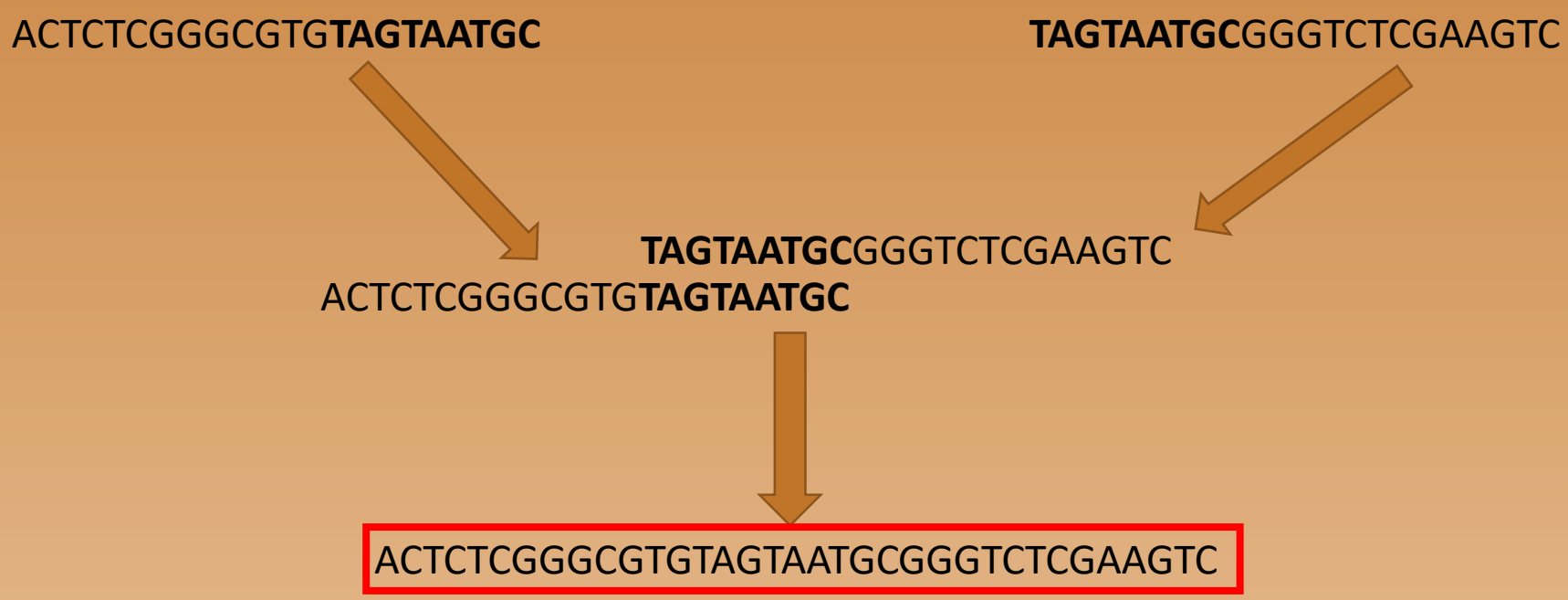
TAGTAATGCGGGTCTCGAAGTC

Assemblers assemble by overlap

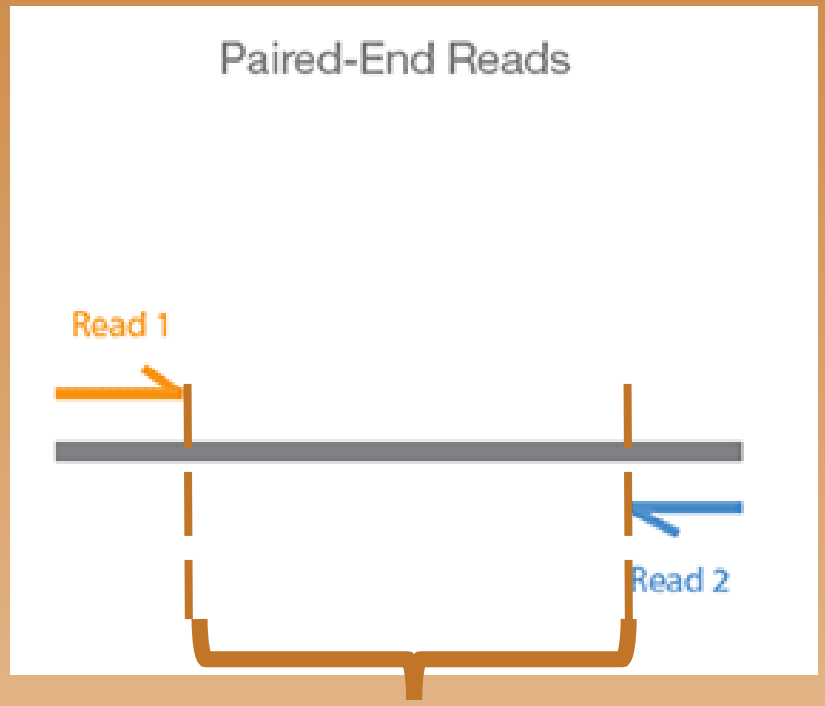
Assemblers assemble by overlap



Assemblers assemble by overlap



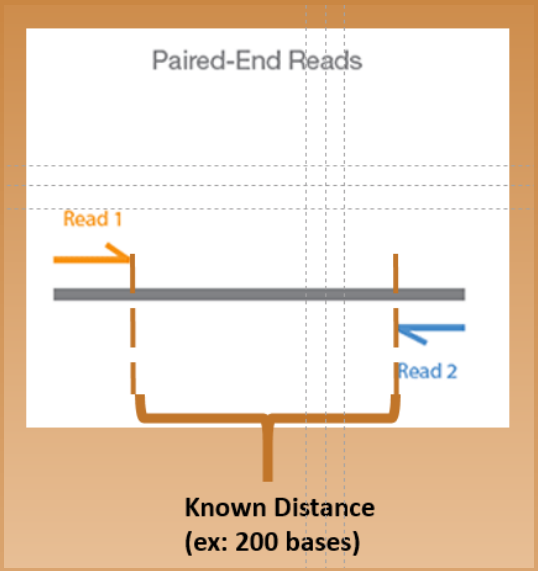
Paired-End reads are a known distance from each other



Known Distance
(ex: 200 bases)

Image Credit: Illumina. *Paired-End Sequencing and Alignment*.

Assemblers take advantage of paired-end reads



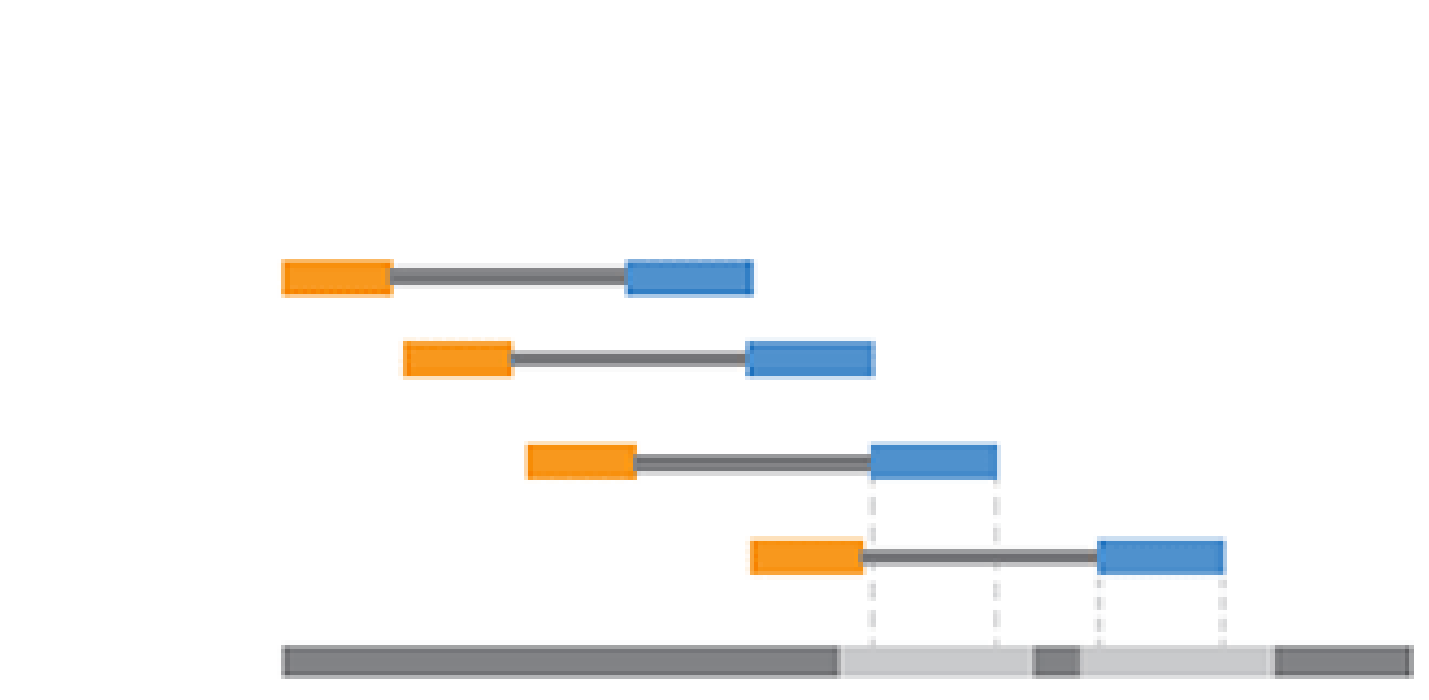


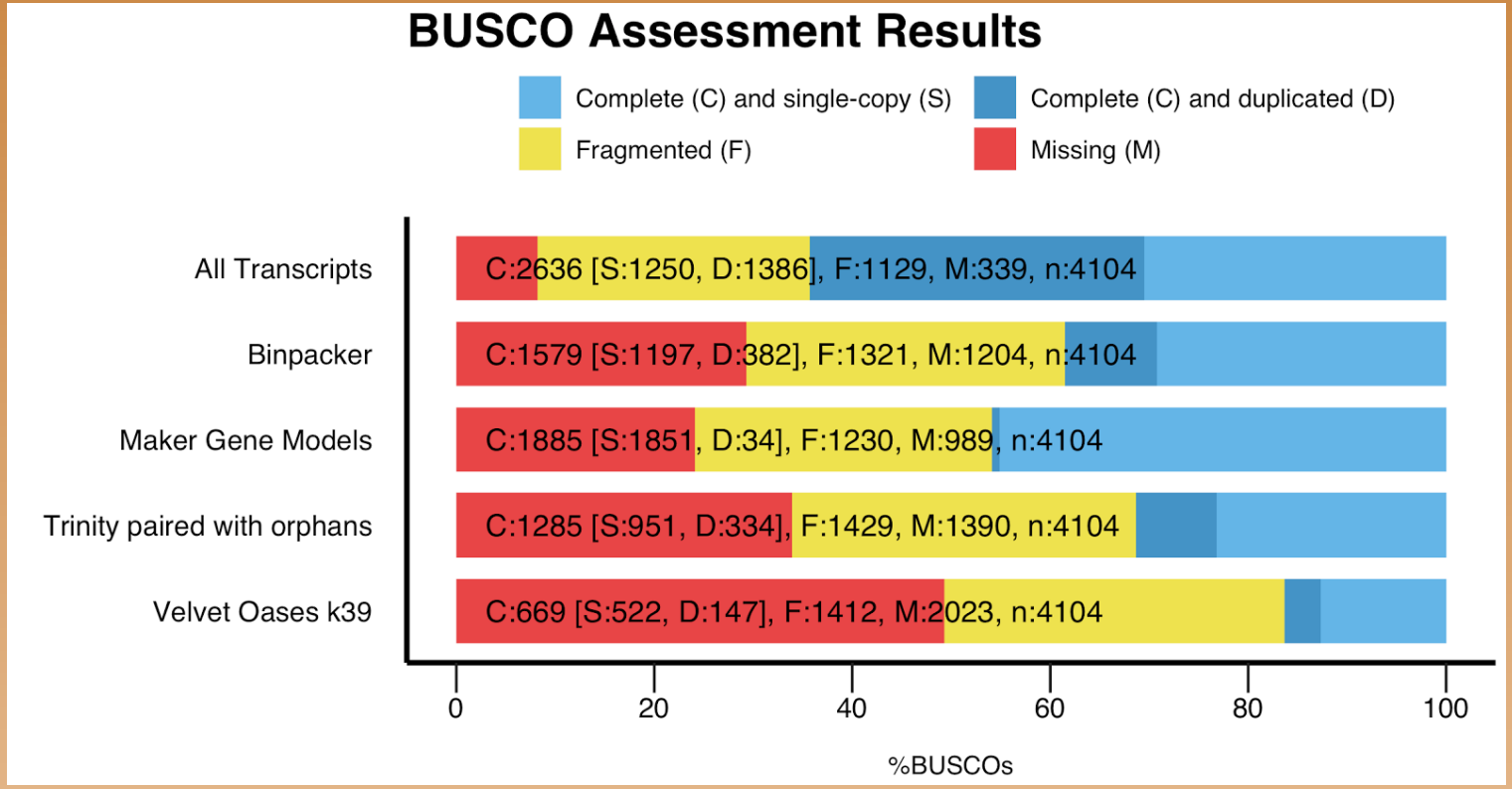
Image Credit: Illumina. *Paired-End Sequencing and Alignment*.

Many programs were used to construct transcriptome assemblies

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

Image Credit: Illumina. *Paired-End Sequencing and Alignment*.

Our assemblies complement each other



BUSCO: Simão et al (2015)

Merged Transcriptome

Merged Transcriptome

86714 contigs

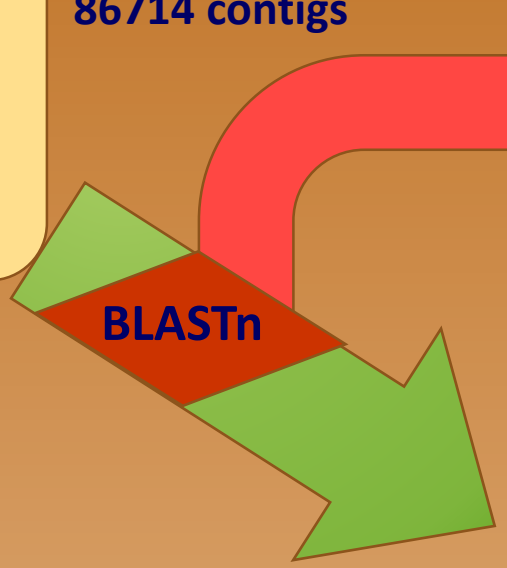
Merged Transcriptome

86714 contigs

Novel lncRNA

Merged Transcriptome

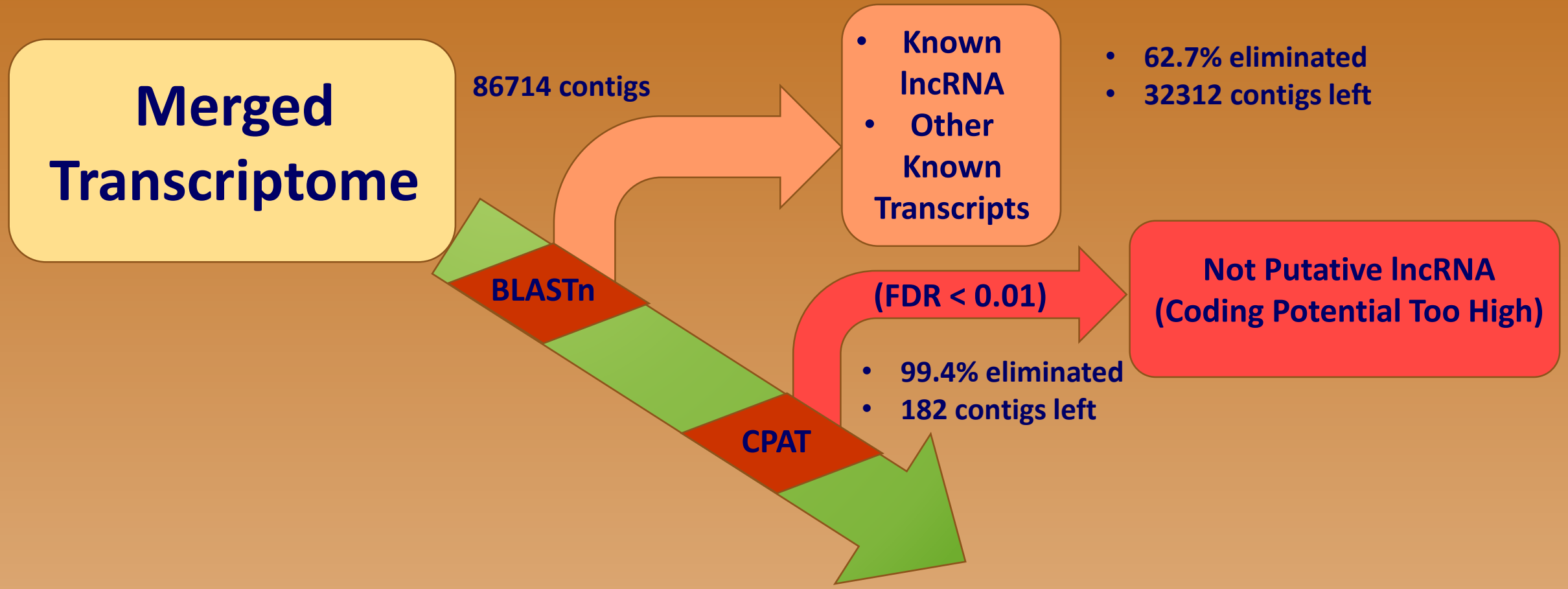
86714 contigs

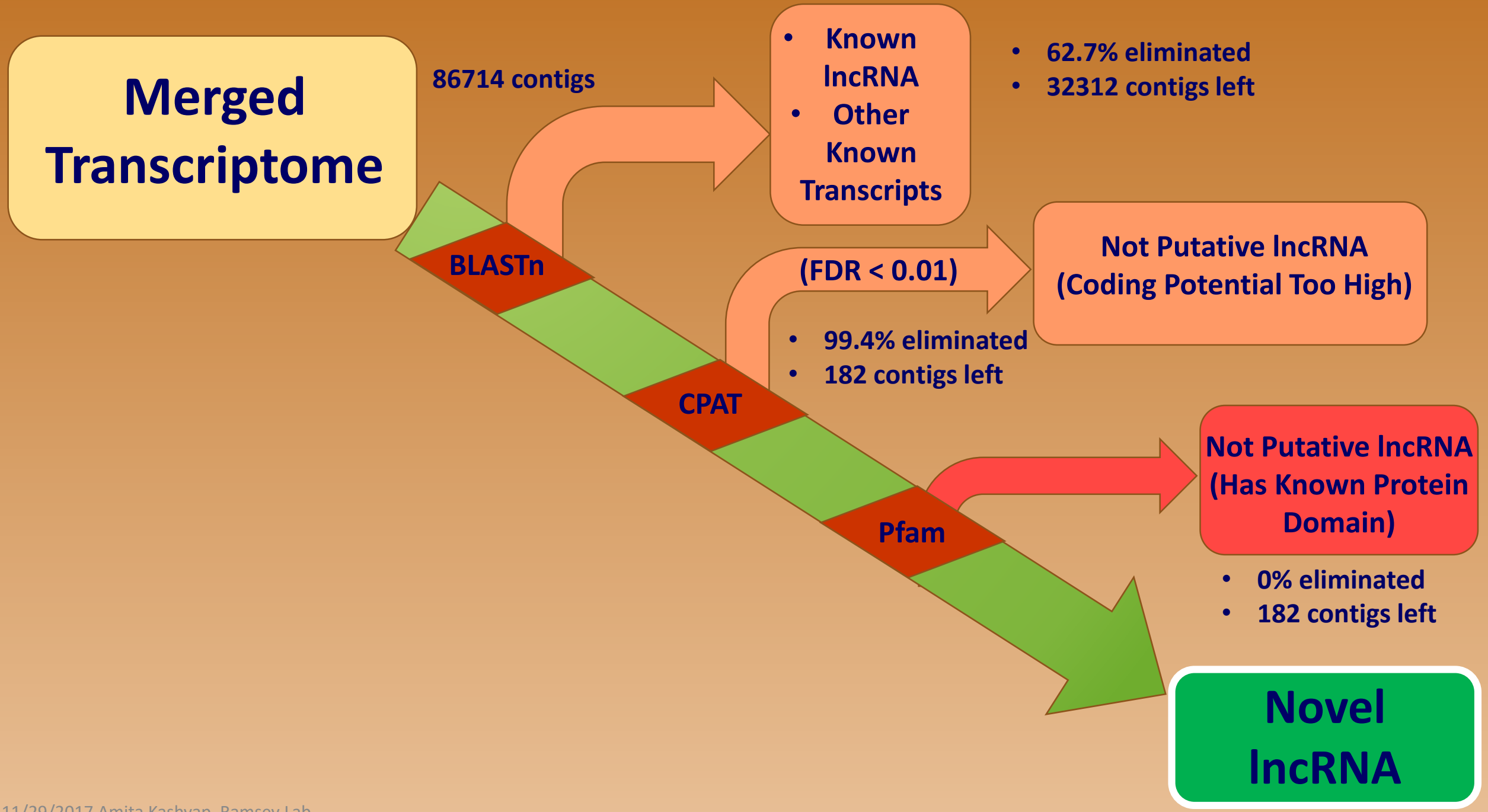


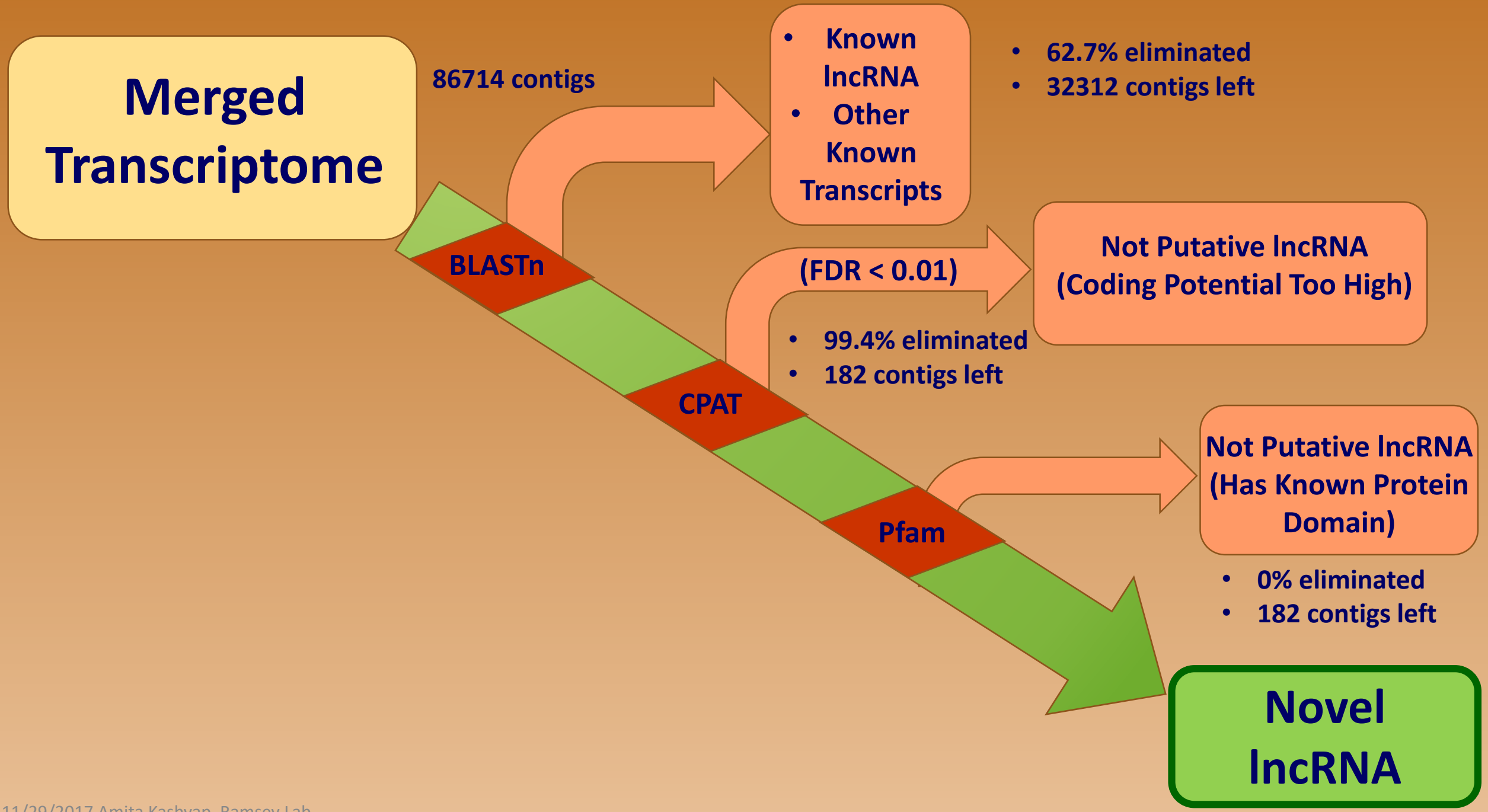
- Known lncRNA
- Other Known Transcripts

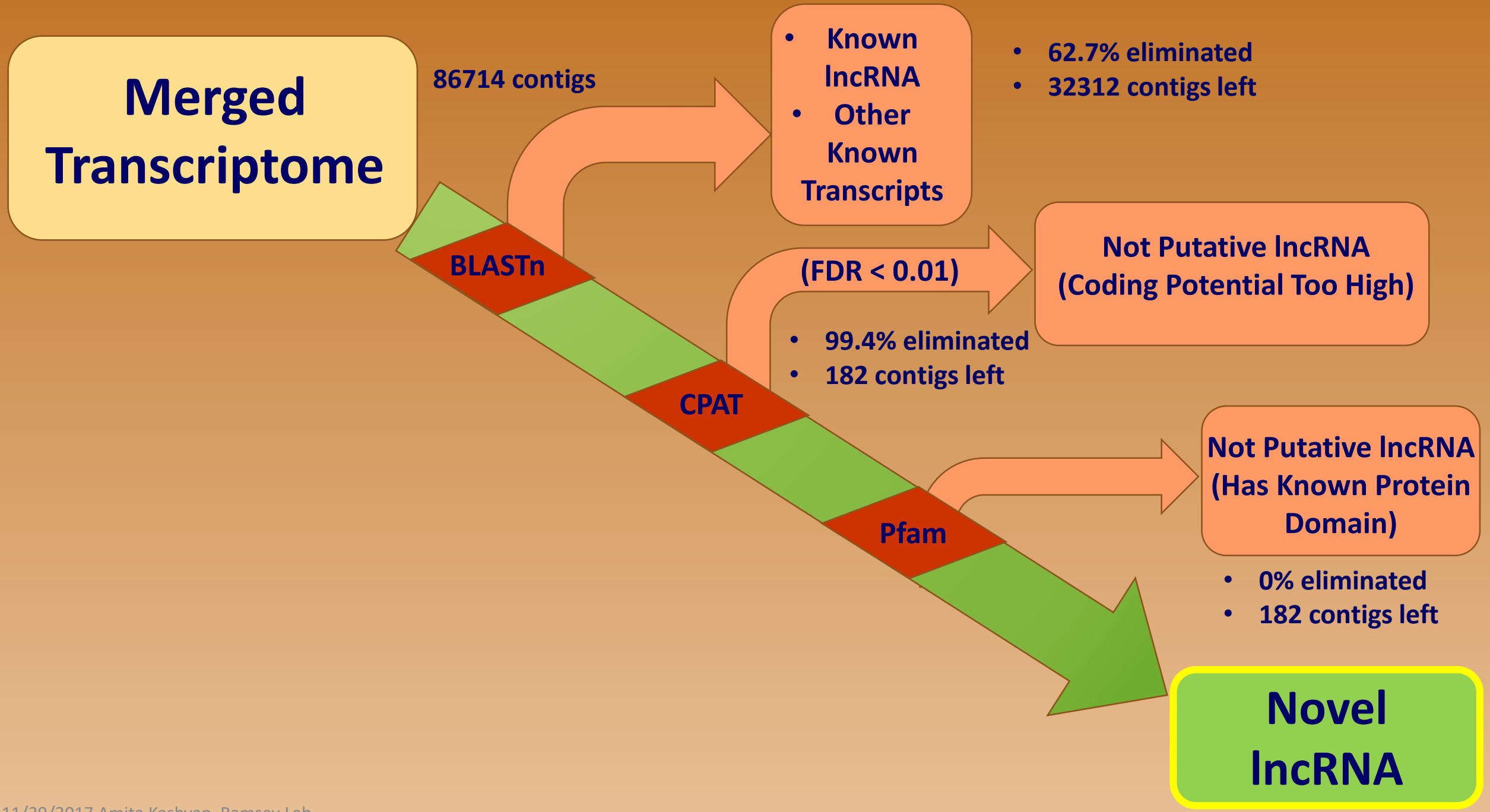
- 62.7% eliminated
- 32312 contigs left

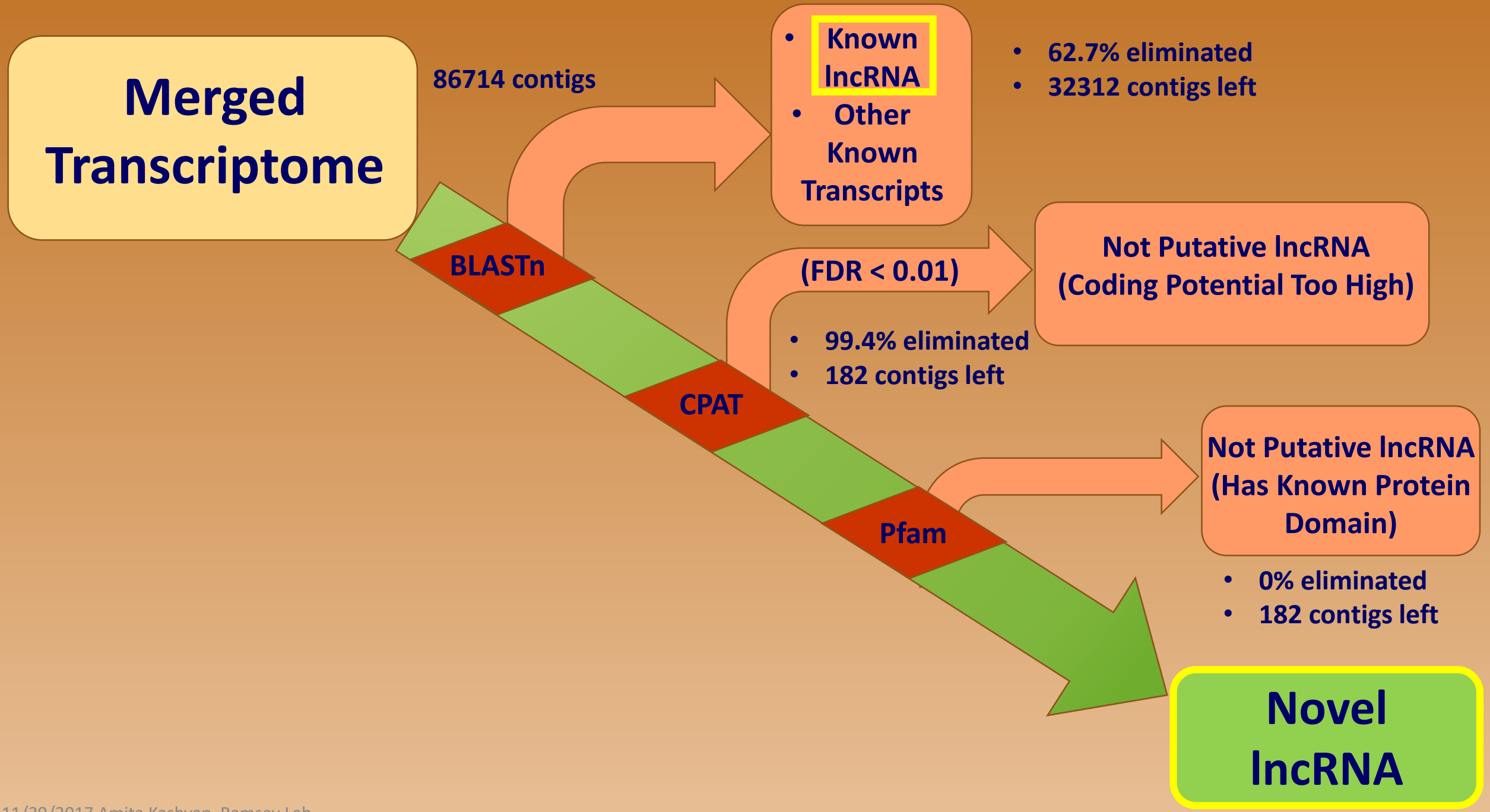
Novel lncRNA


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
Novel lncRNA


Merged Transcriptome
86714 contigs
BLASTn
Known lncRNA
Other Known Transcripts
62.7% eliminated
32312 contigs left
(FDR < 0.01)
CPAT
Not Putative lncRNA (Coding Potential Too High)
99.4% eliminated
182 contigs left
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)
CPAT
Not Putative lncRNA (Coding Potential Too High)
99.4% eliminated
182 contigs left
Pfam
Not Putative lncRNA (Has Known Protein Domain)
0% eliminated
182 contigs left
Novel lncRNA

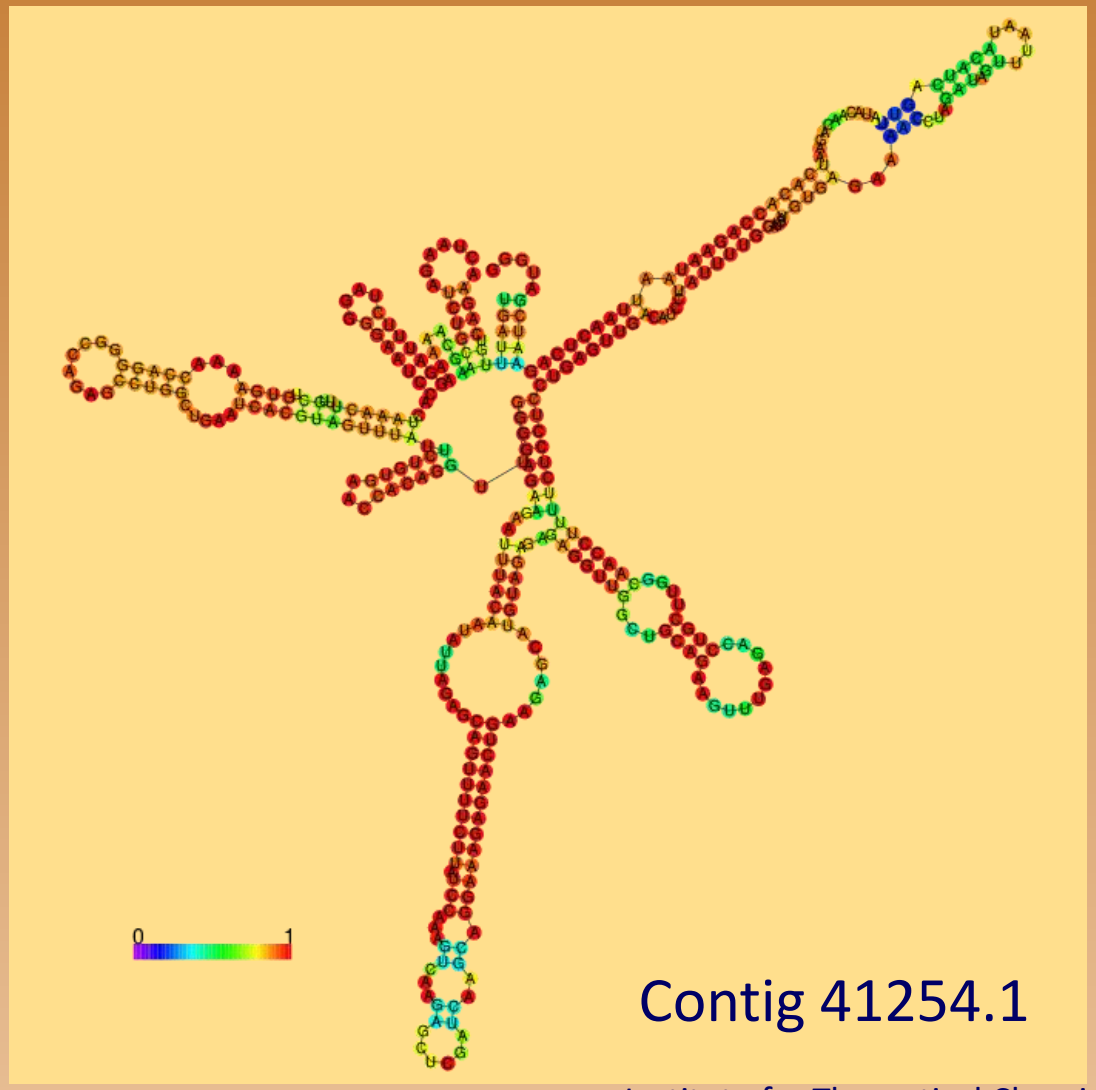

Merged Transcriptome
86714 contigs
BLASTn
Known lncRNA
Other Known Transcripts
62.7% eliminated
32312 contigs left
(FDR < 0.01)
CPAT
Not Putative lncRNA (Coding Potential Too High)
99.4% eliminated
182 contigs left
Pfam
Not Putative lncRNA (Has Known Protein Domain)
0% eliminated
182 contigs left
Novel lncRNA


Merged Transcriptome
86714 contigs
BLASTn
Known lncRNA
Other Known Transcripts
62.7% eliminated
32312 contigs left
(FDR < 0.01)
CPAT
Not Putative lncRNA (Coding Potential Too High)
99.4% eliminated
182 contigs left
Pfam
Not Putative lncRNA (Has Known Protein Domain)
0% eliminated
182 contigs left
Novel lncRNA

Overview

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

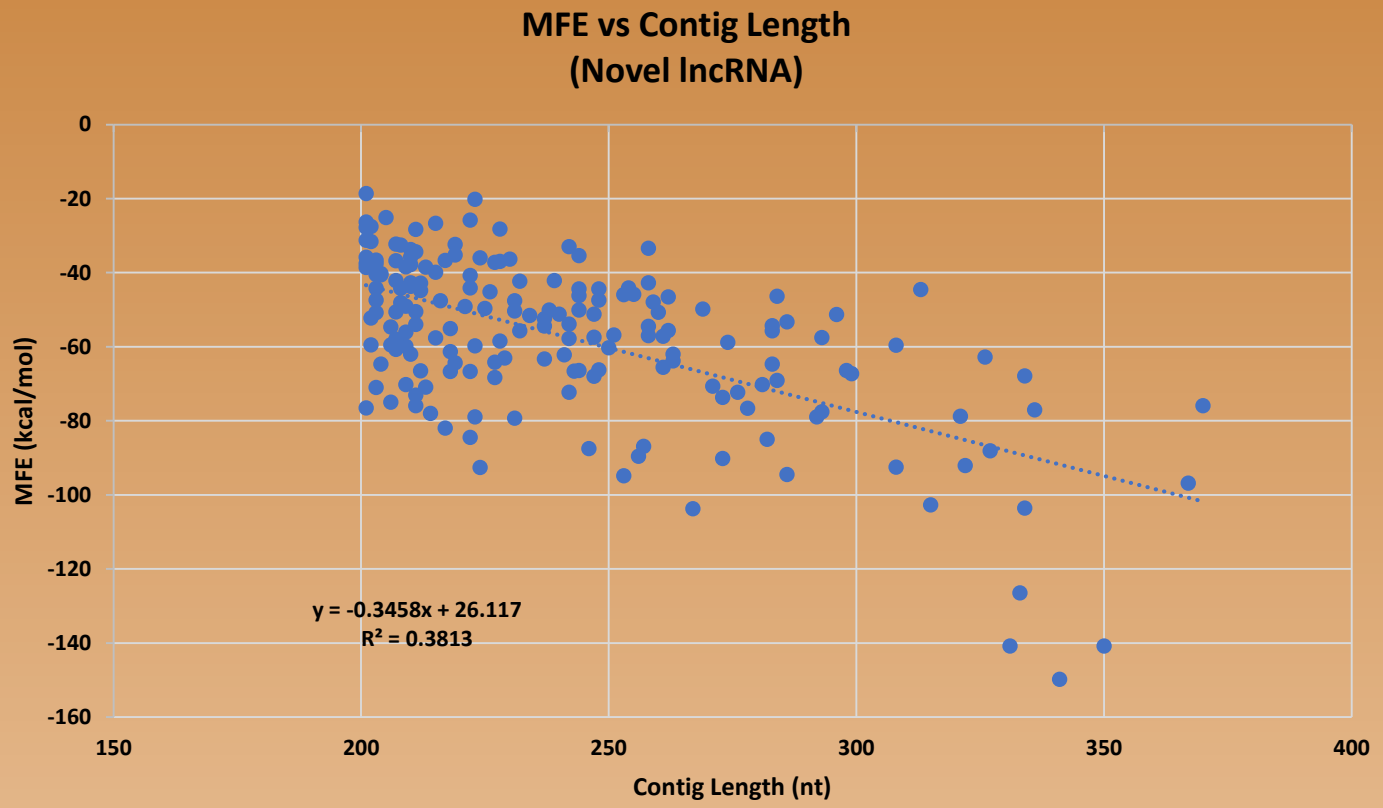
Secondary Structure Analysis



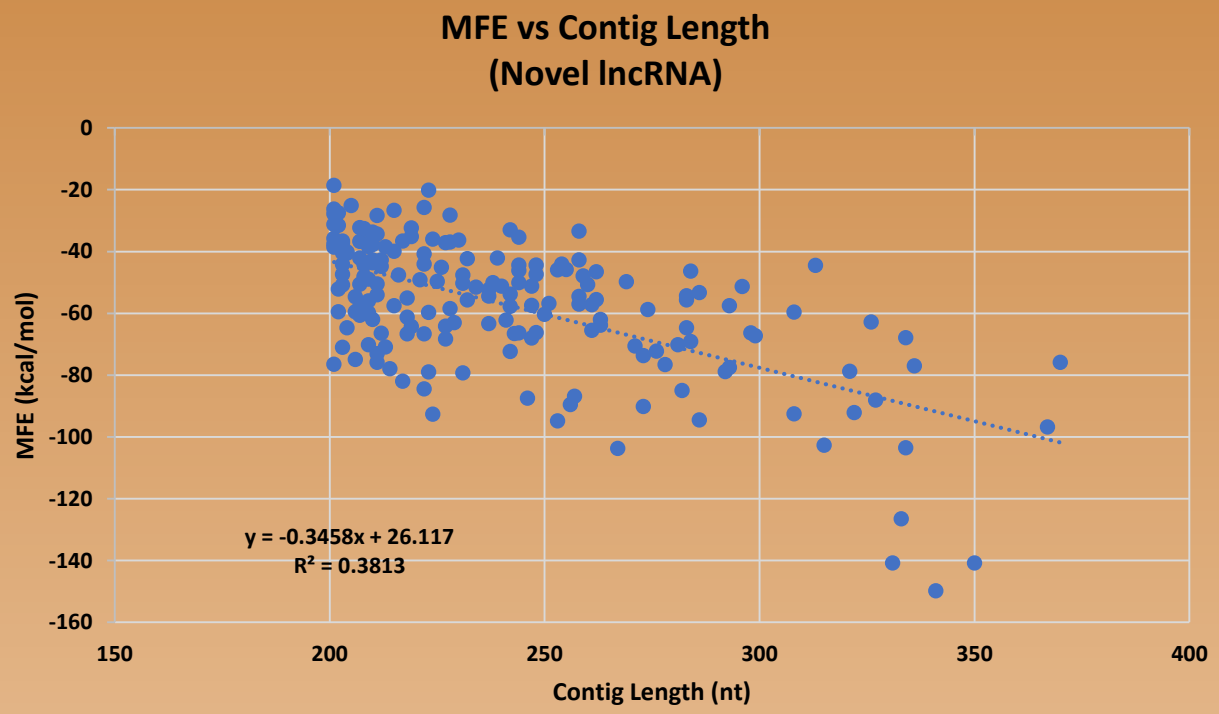
Contig 41254.1

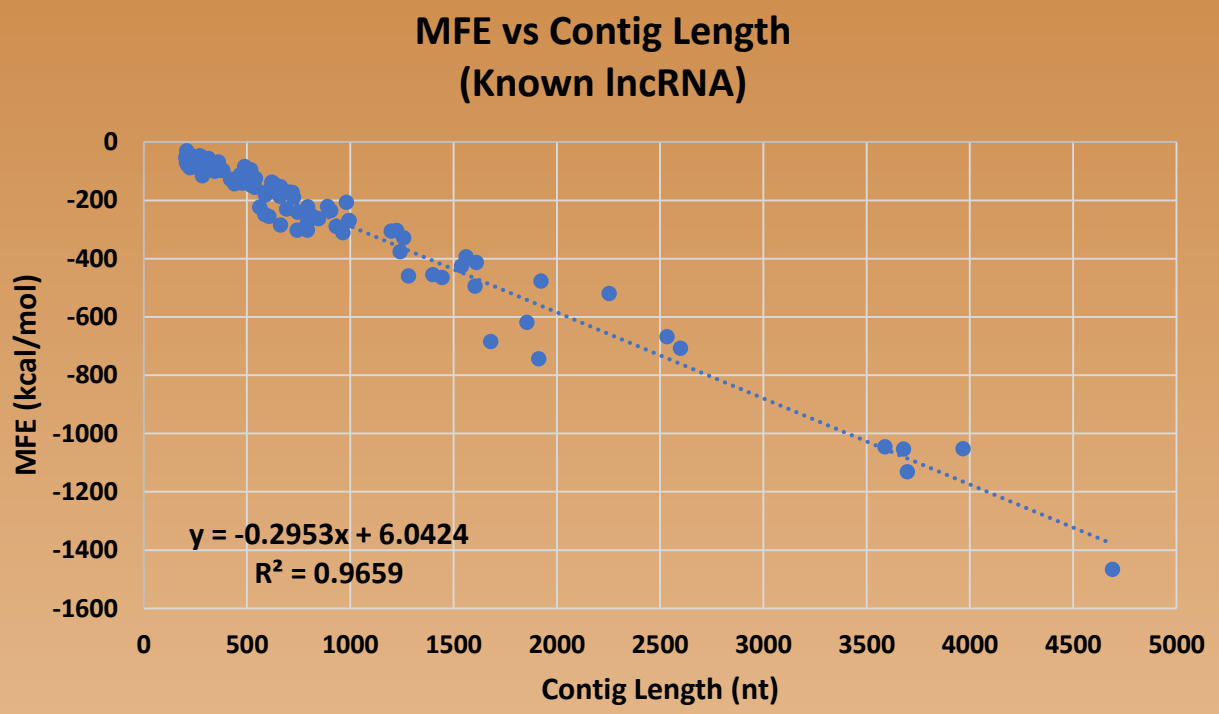
Institute for Theoretical Chemistry, University of Vienna. RNAfold WebServer. http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi

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

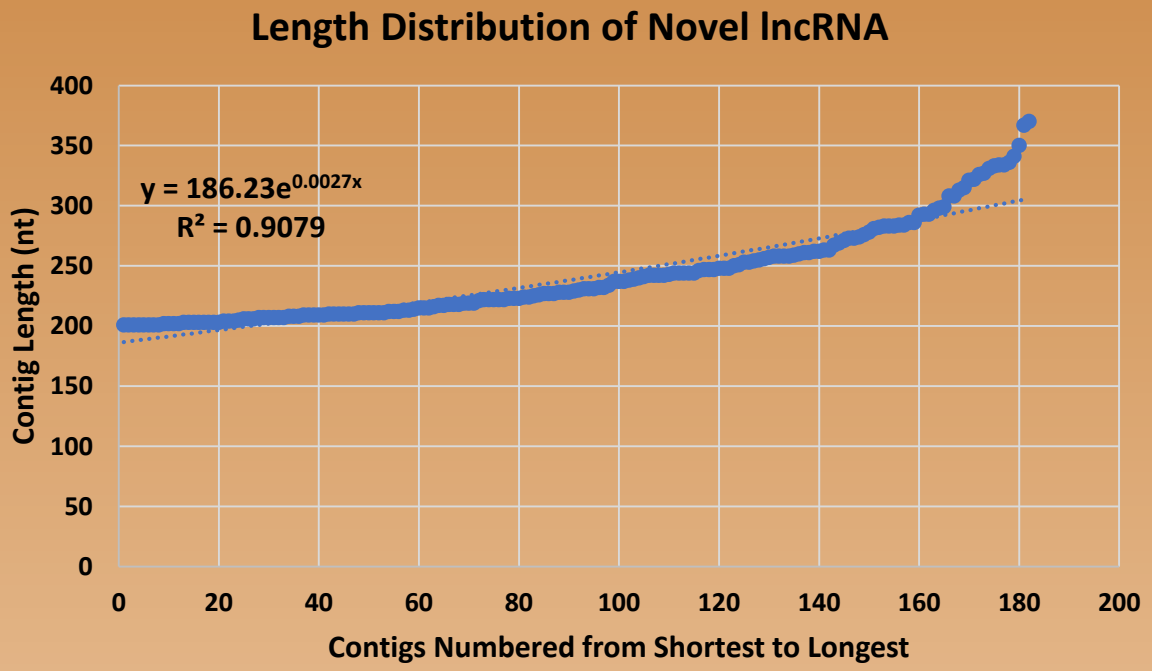


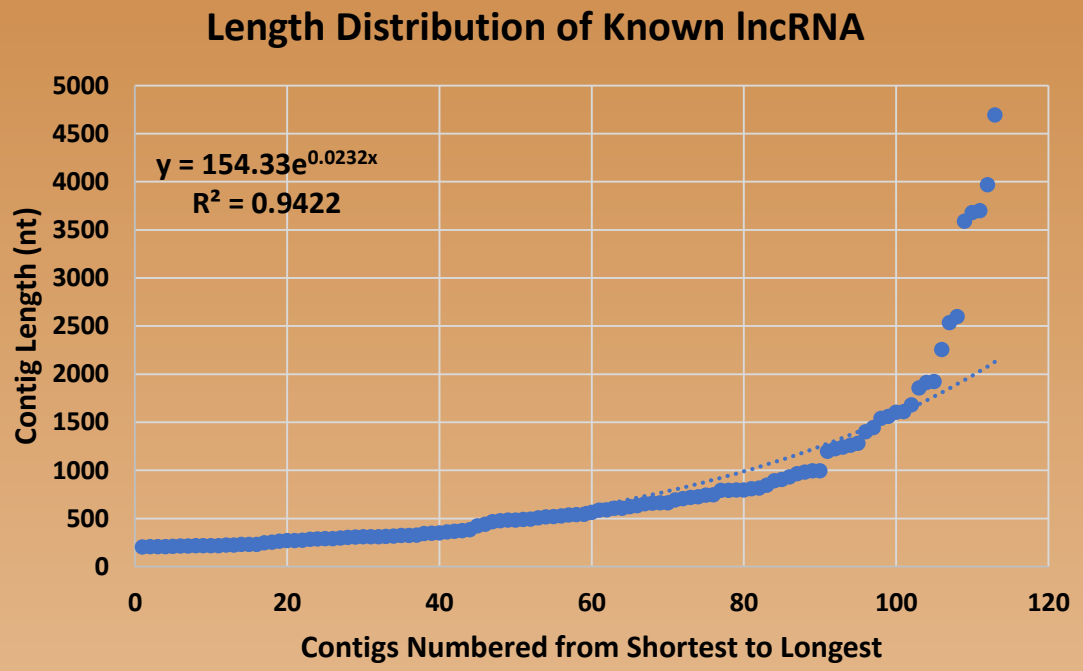
The novel lncRNA have similar characteristics to the known lncRNA



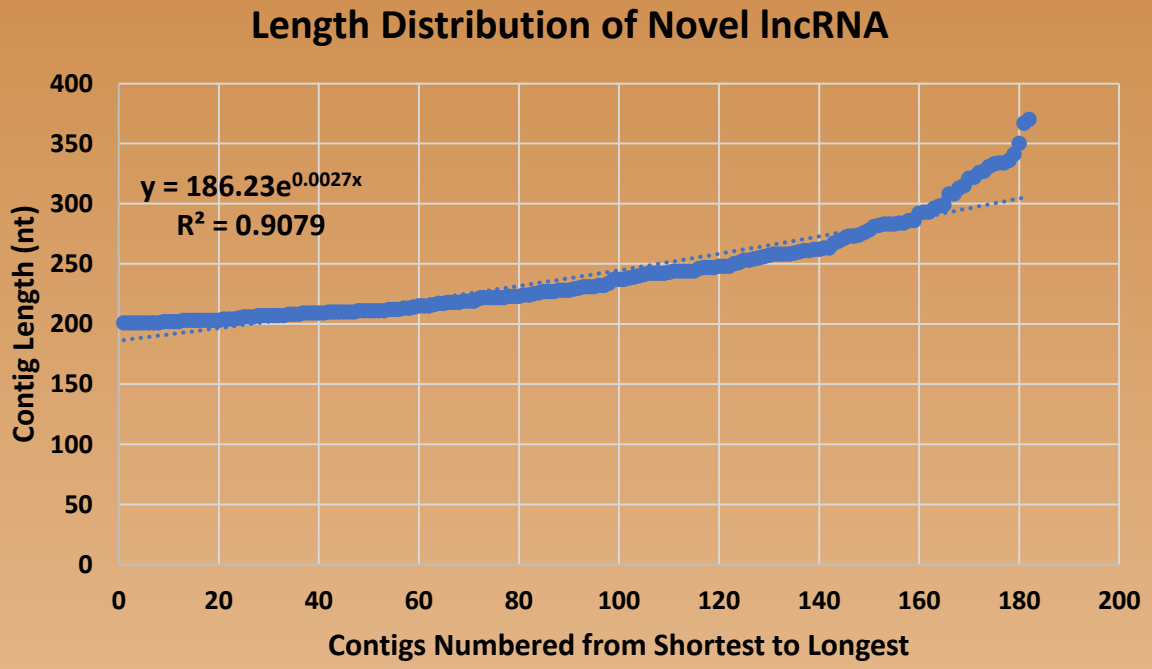


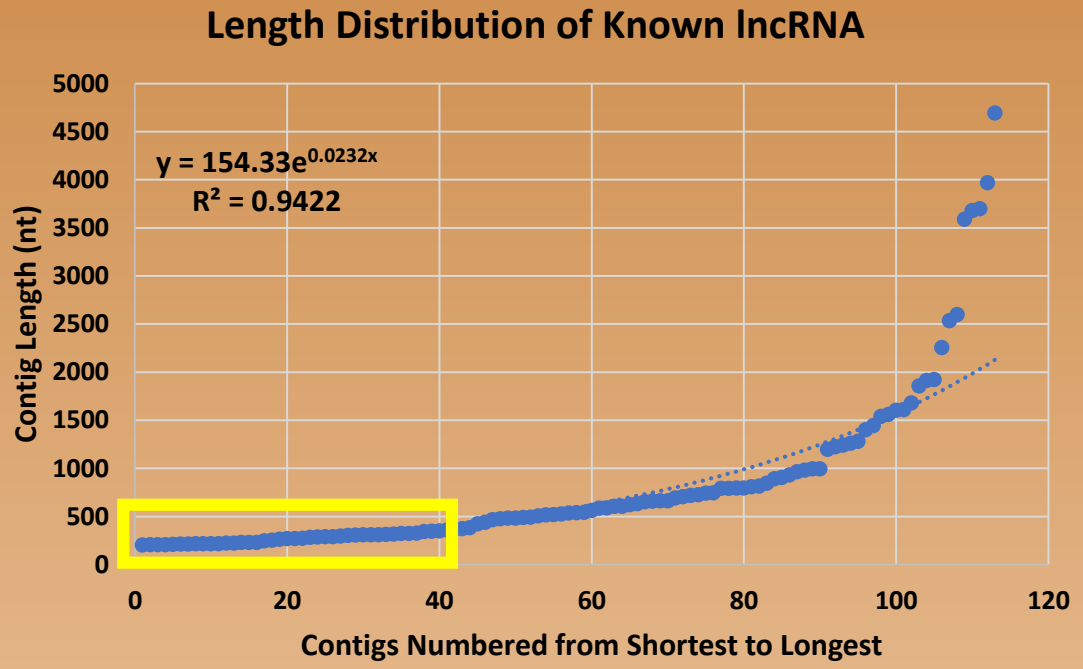
The novel lncRNA represent the lower 40th percentile by length of the known lncRNA



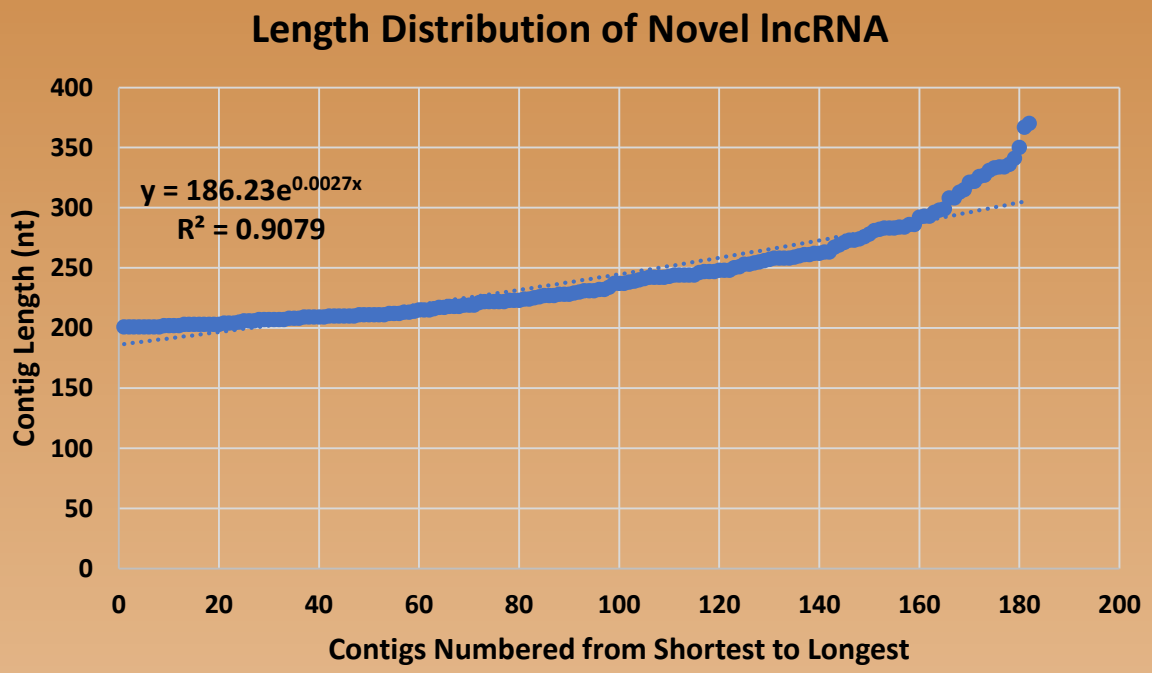


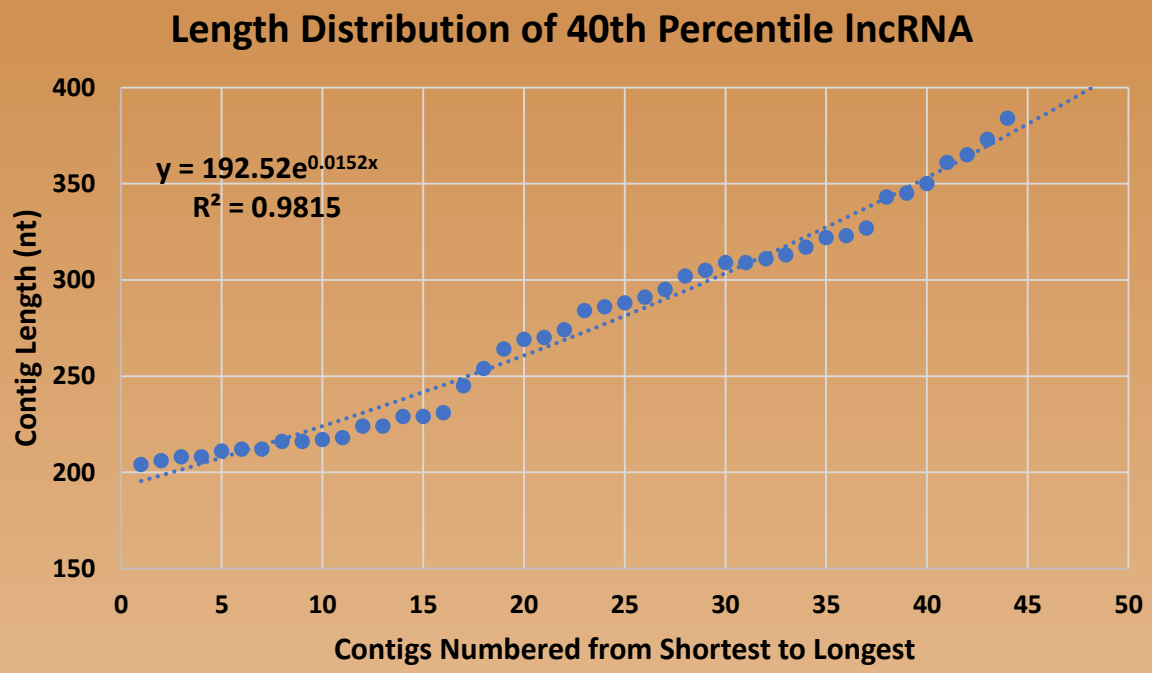
The novel lncRNA represent the lower 40th percentile by length of the known lncRNA



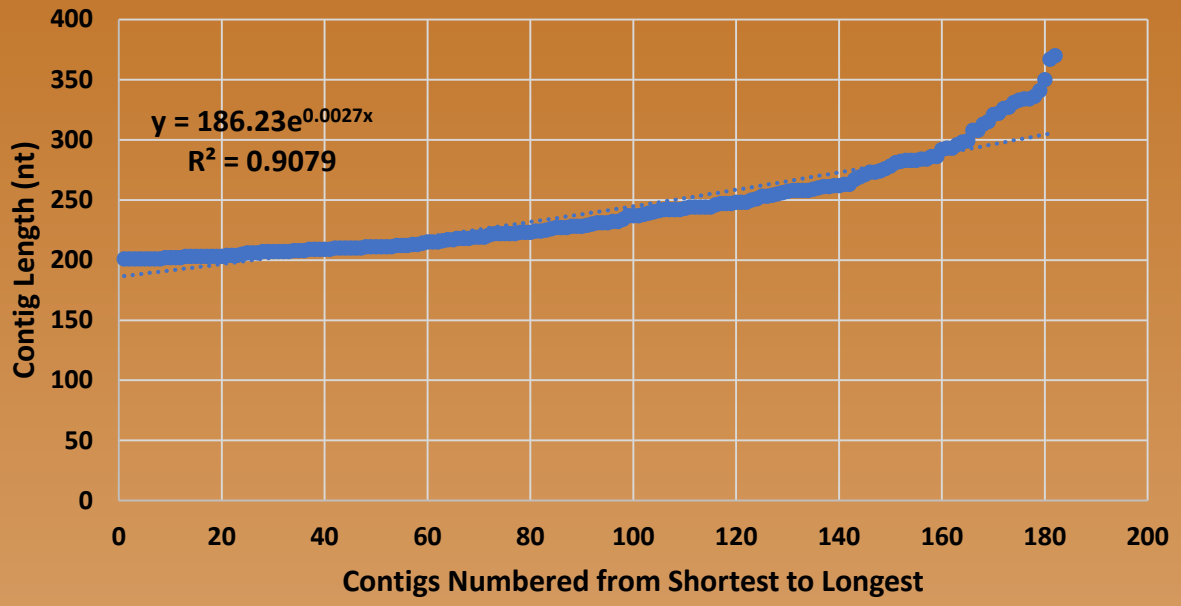


The novel lncRNA have similar characteristics to the 40th percentile lncRNA

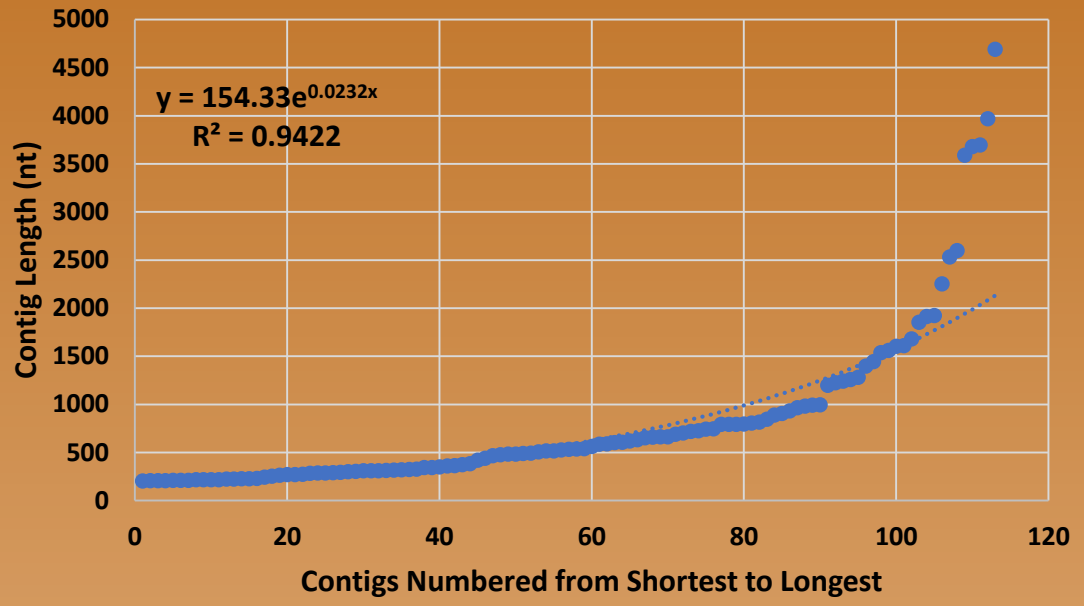




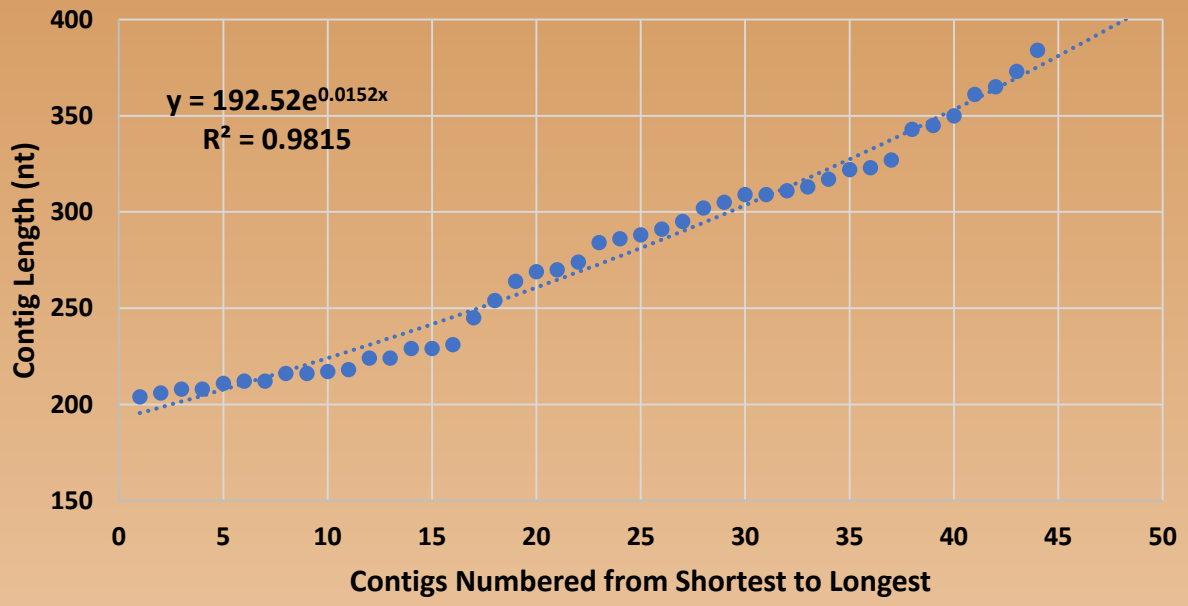
Length Distribution of Novel lncRNA



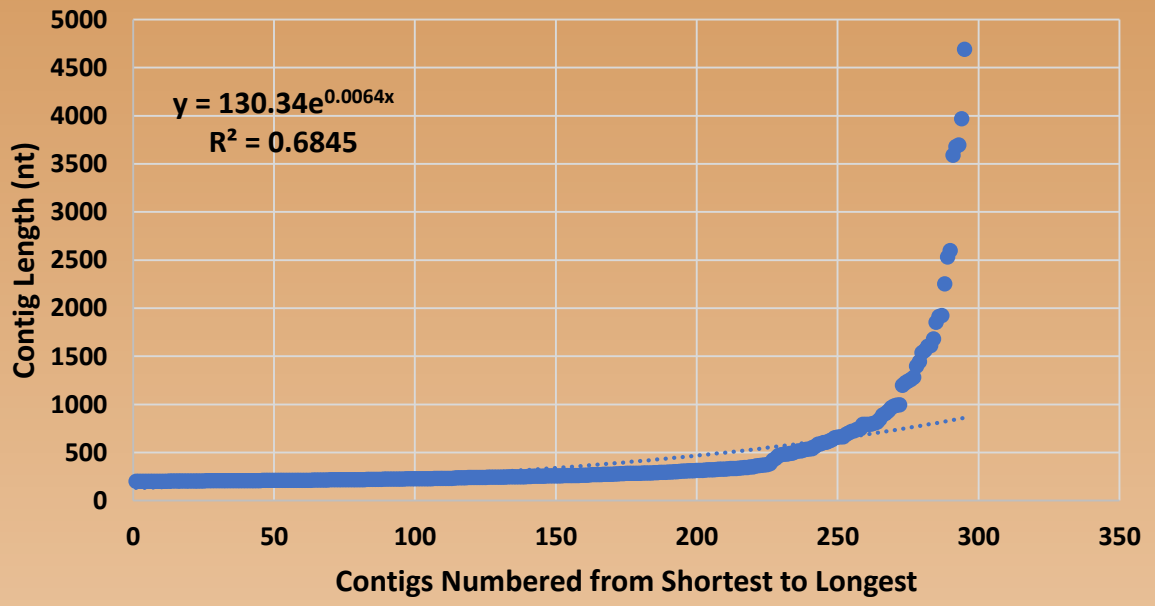
Length Distribution of Known lncRNA



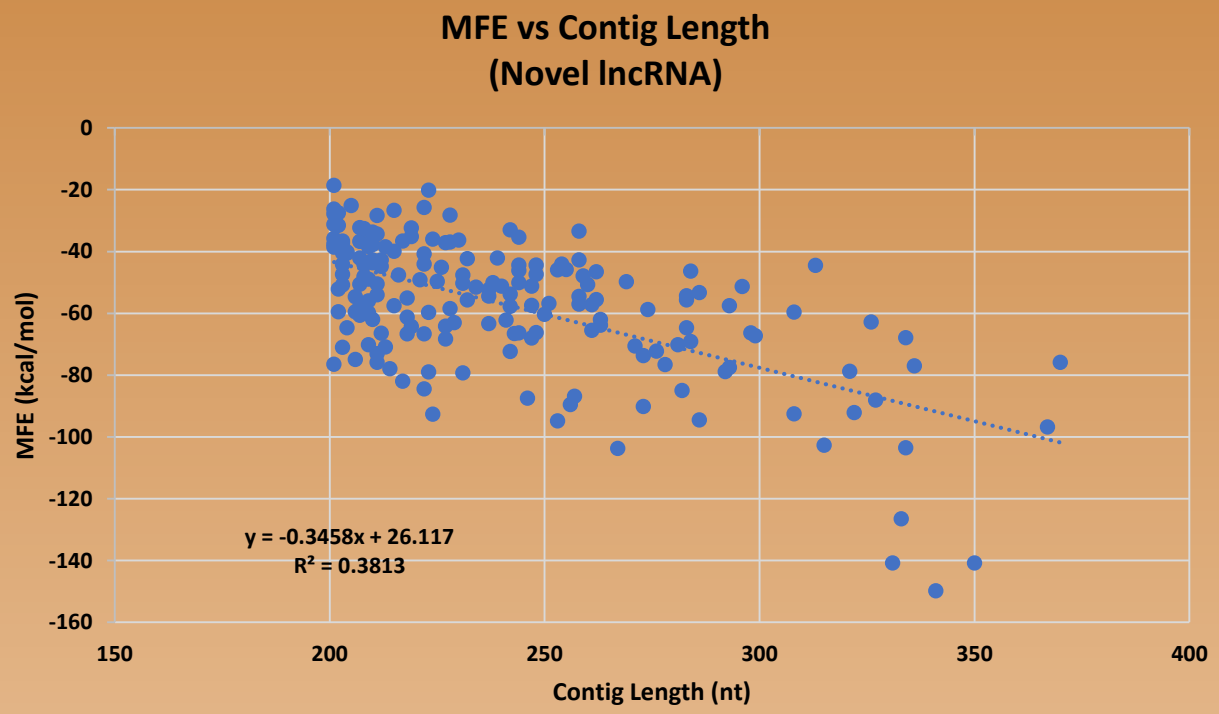
Length Distribution of 40th Percentile lncRNA

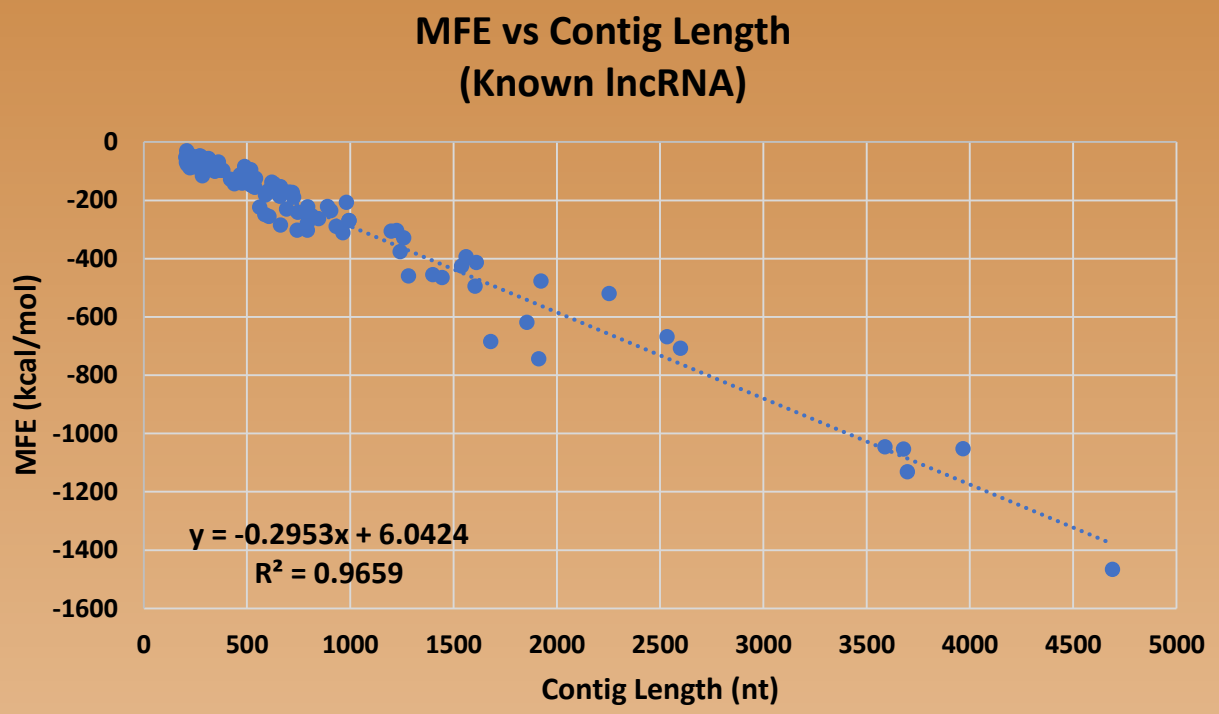


Length Distribution of Pooled lncRNA

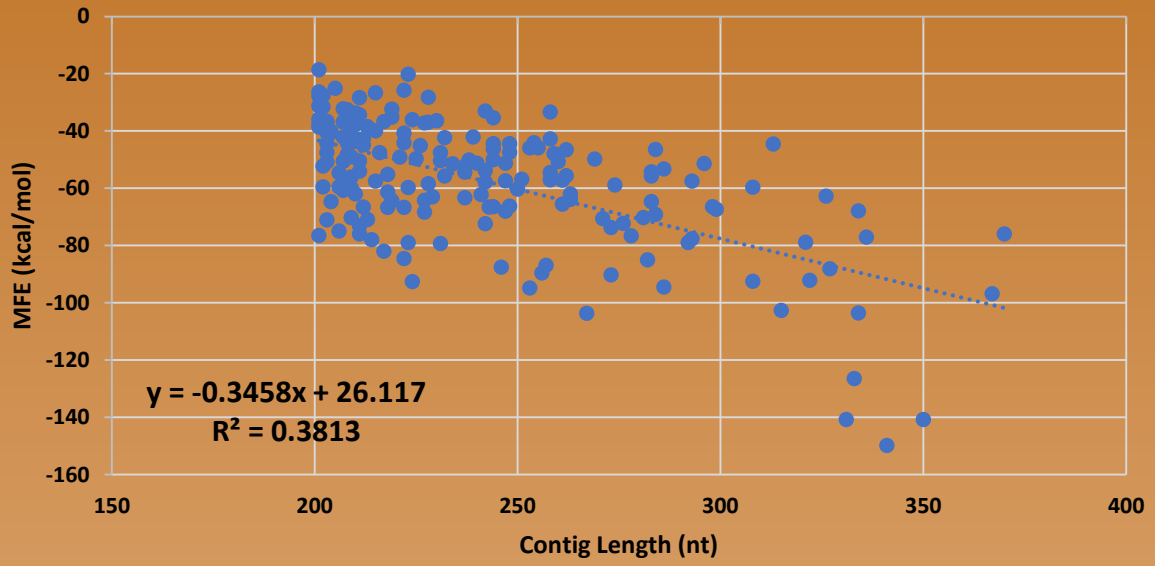


The novel lncRNA have similar characteristics to the known lncRNA

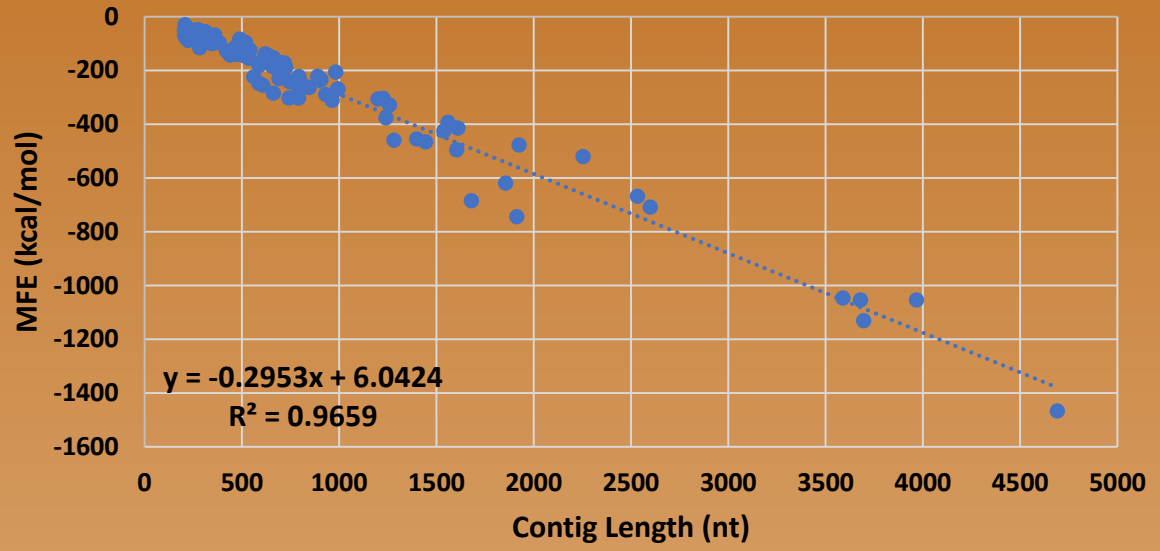




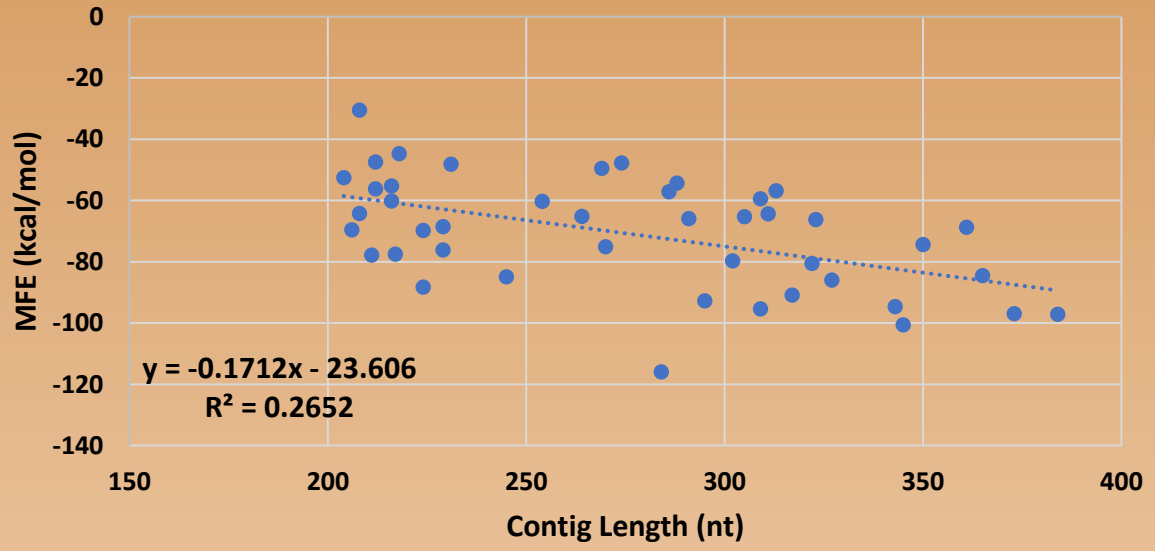
MFE vs Contig Length (Novel lncRNA)



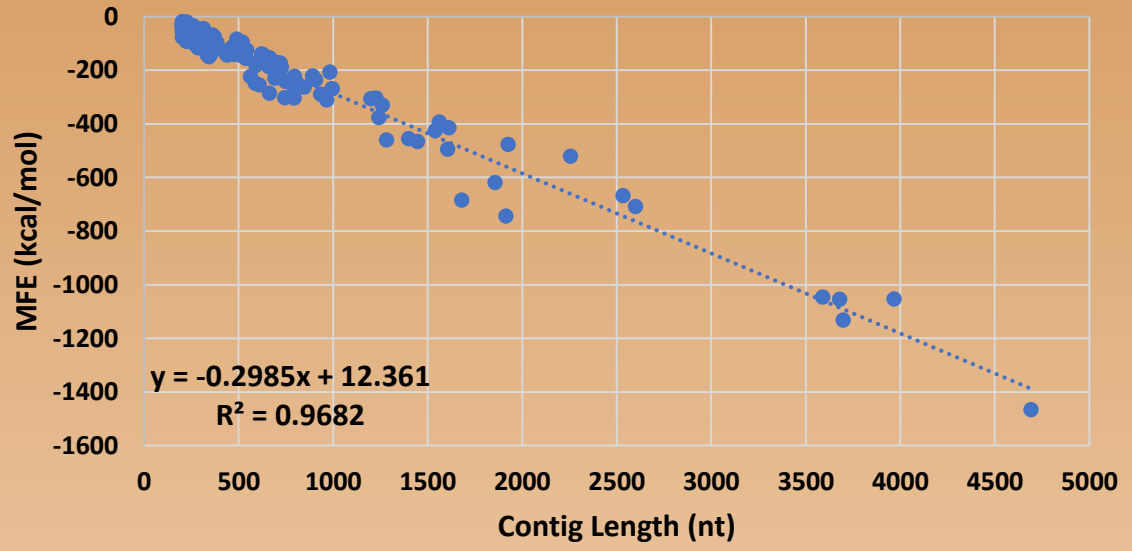
MFE vs Contig Length (Known lncRNA)



MFE vs Contig Length (40th Percentile lncRNA)

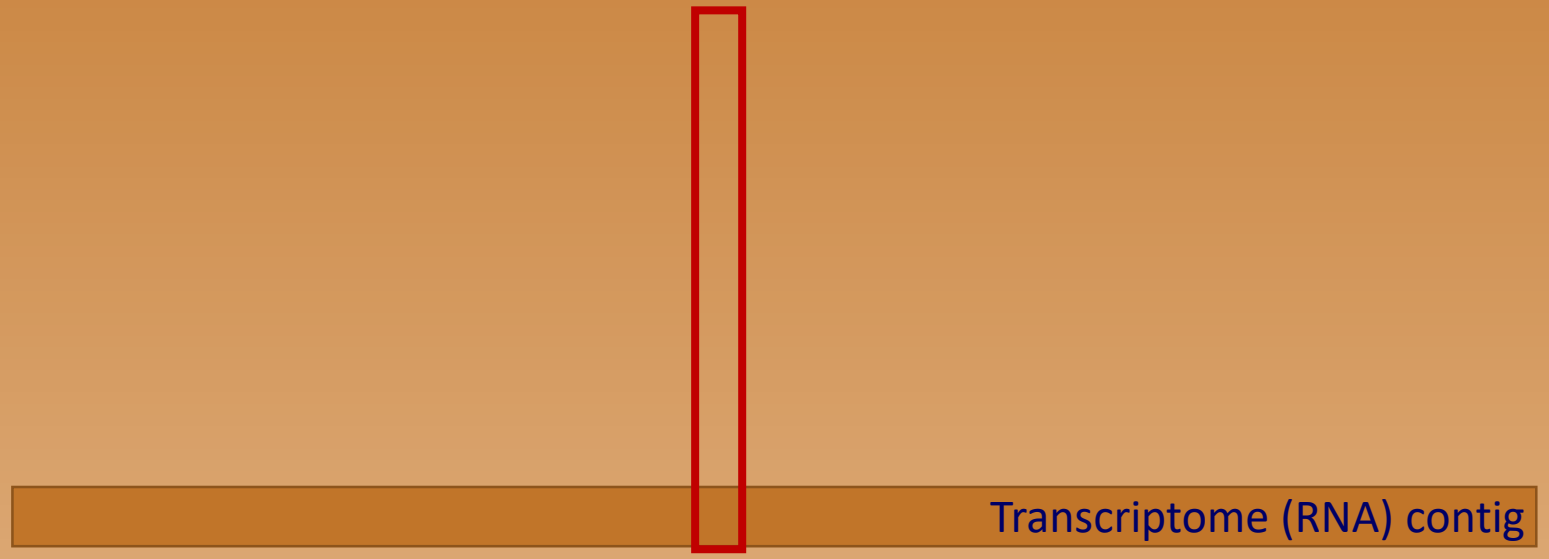


MFE vs Contig Length (Pooled lncRNA)

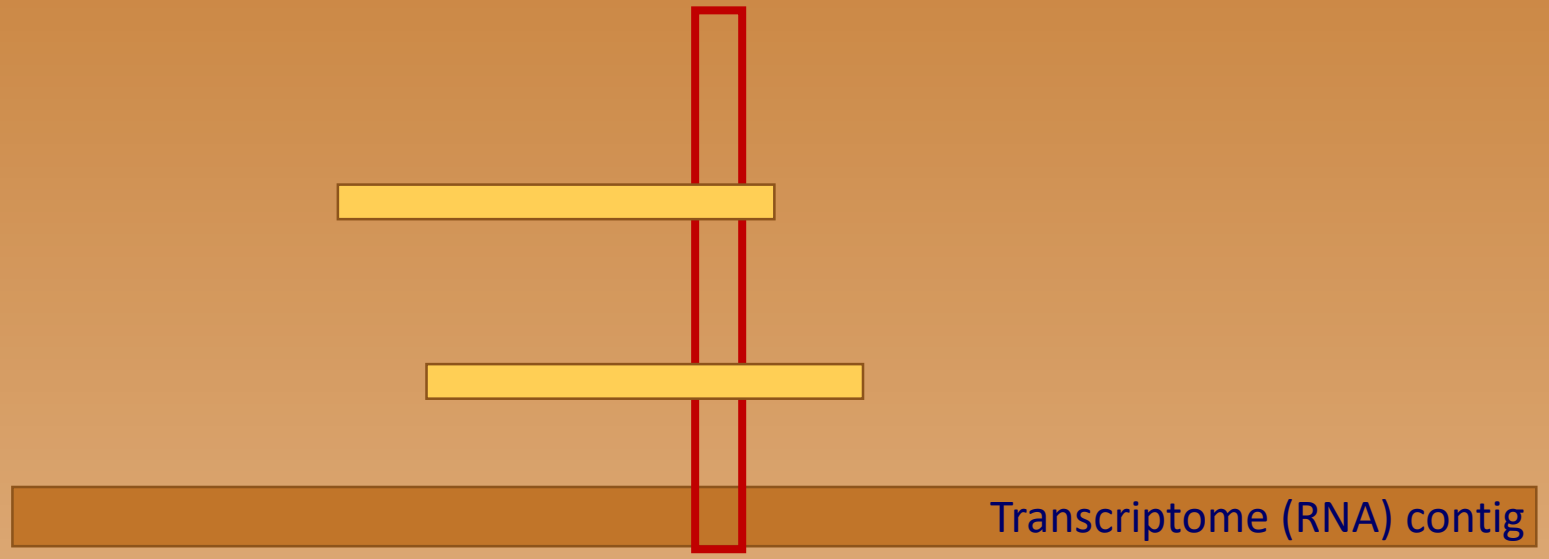


We inspected coverage

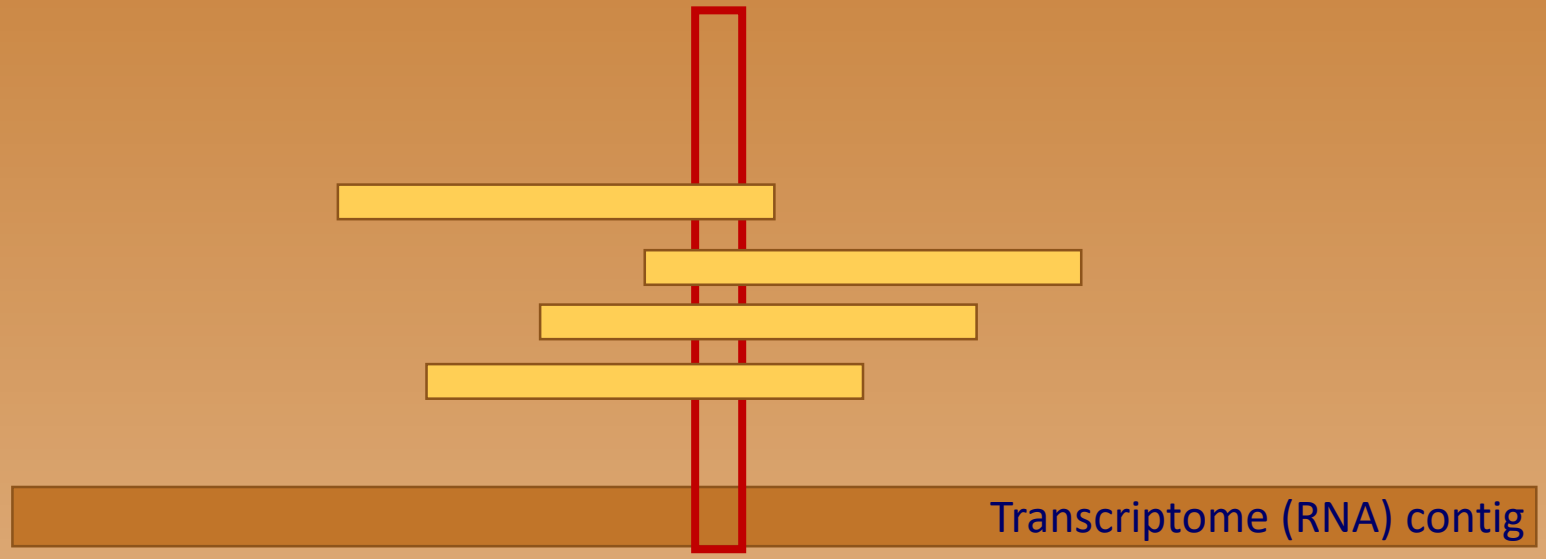
We inspected coverage



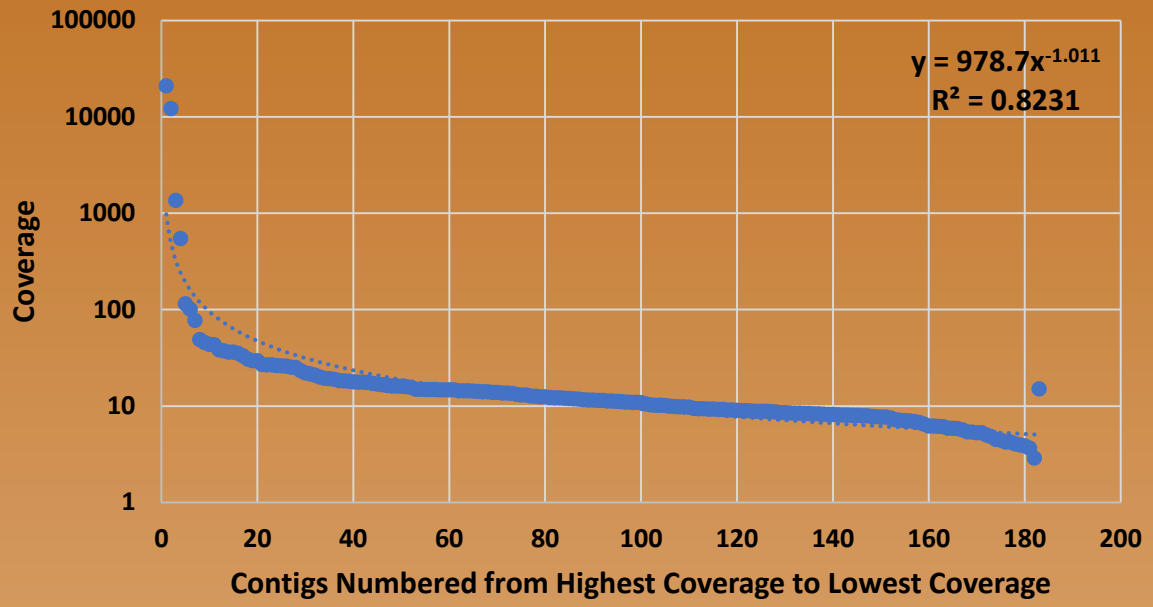
We inspected coverage



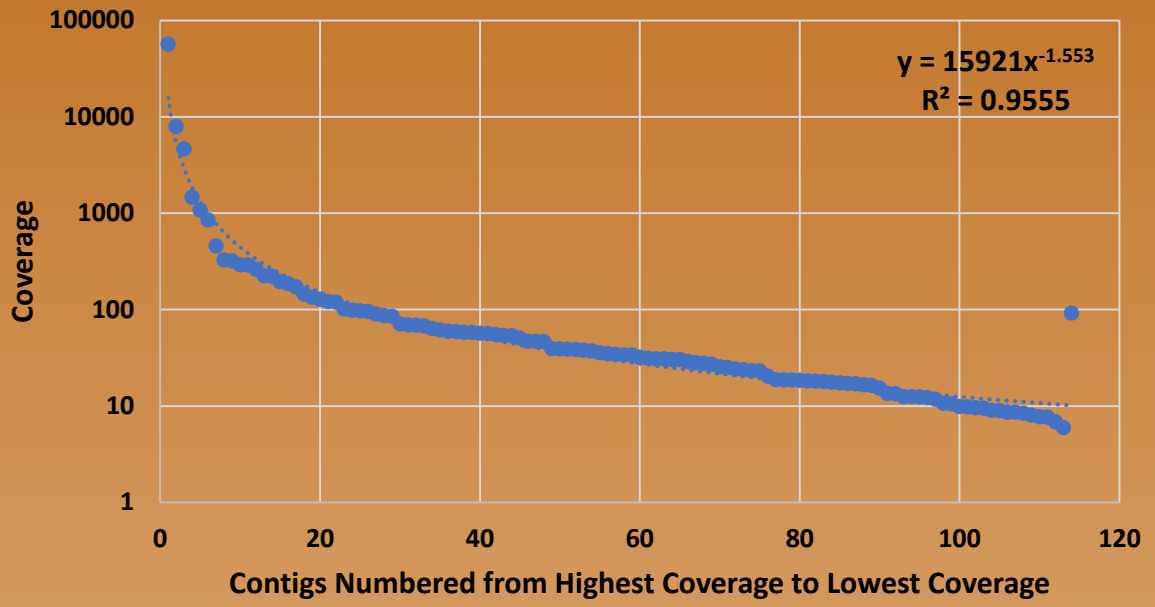
We inspected coverage



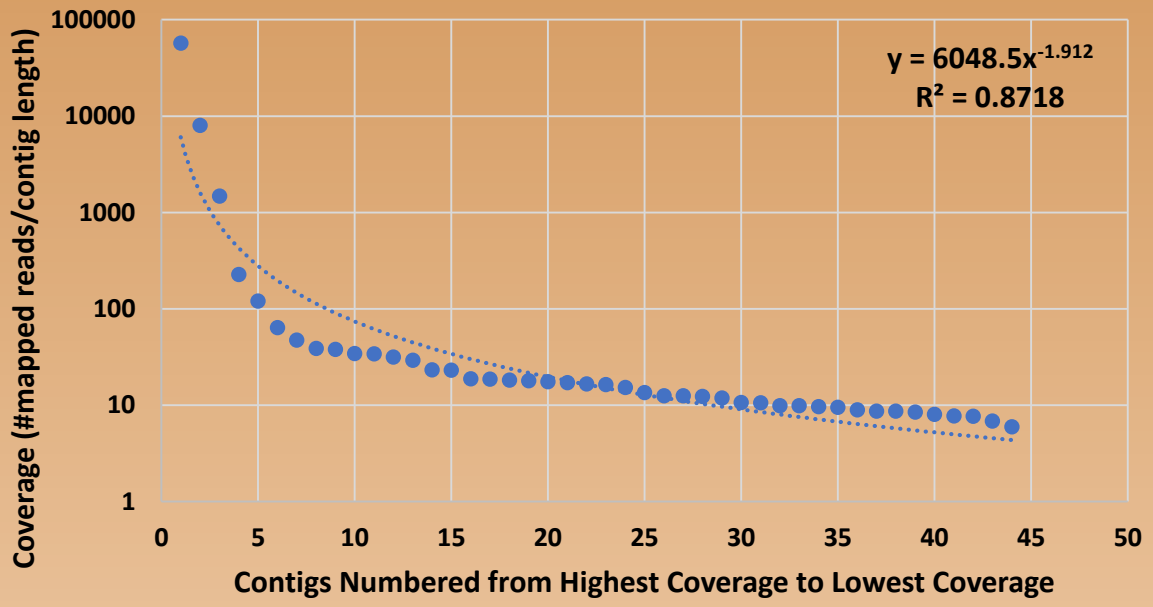
Coverage Distribution of Novel lncRNA



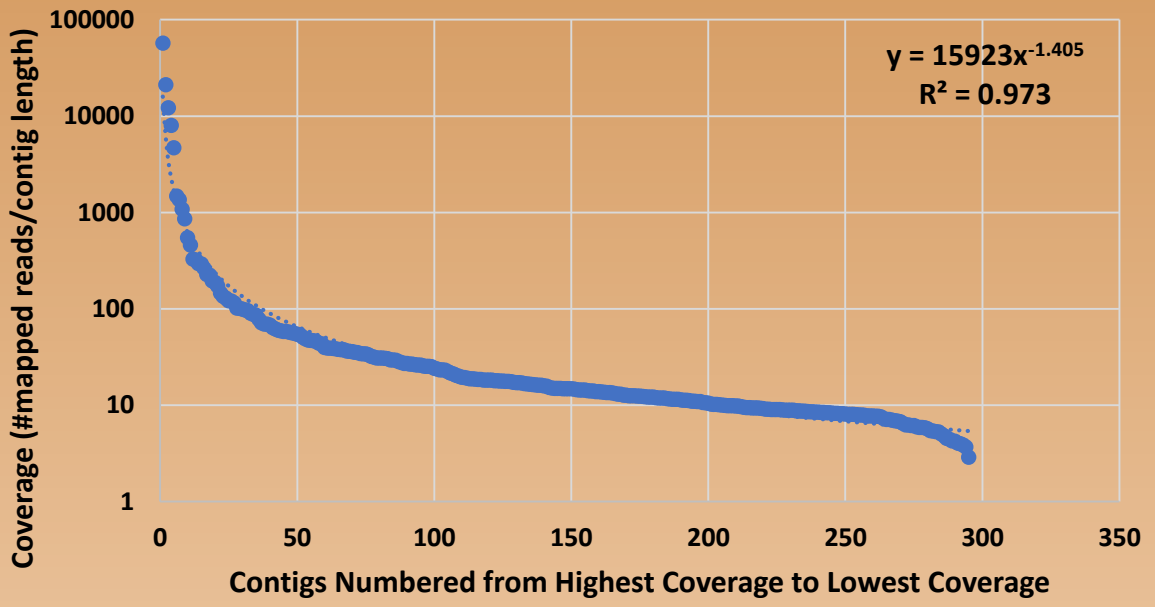
Coverage Distribution of Known lncRNA



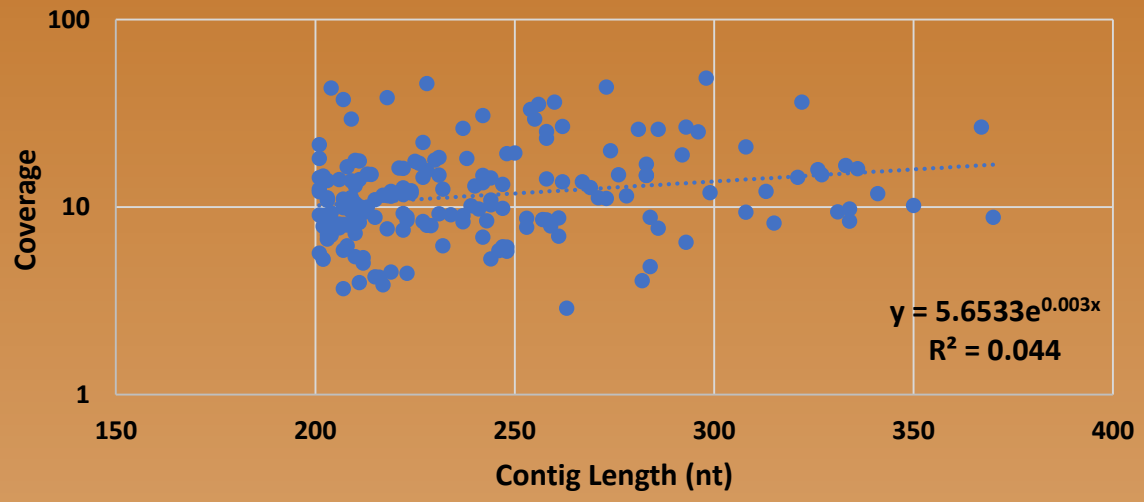
Coverage Distribution of 40th Percentile lncRNA



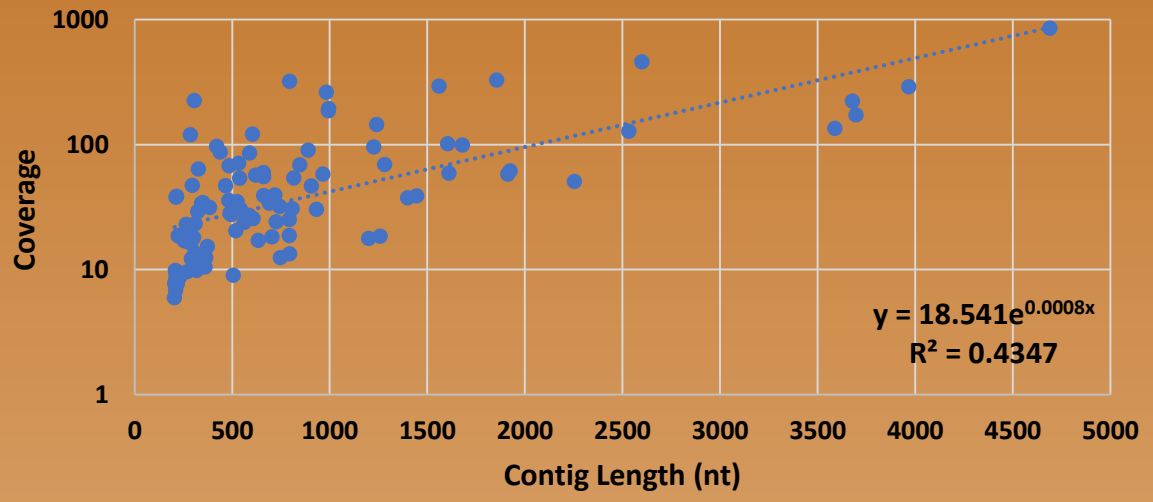
Coverage Distribution of Pooled lncRNA



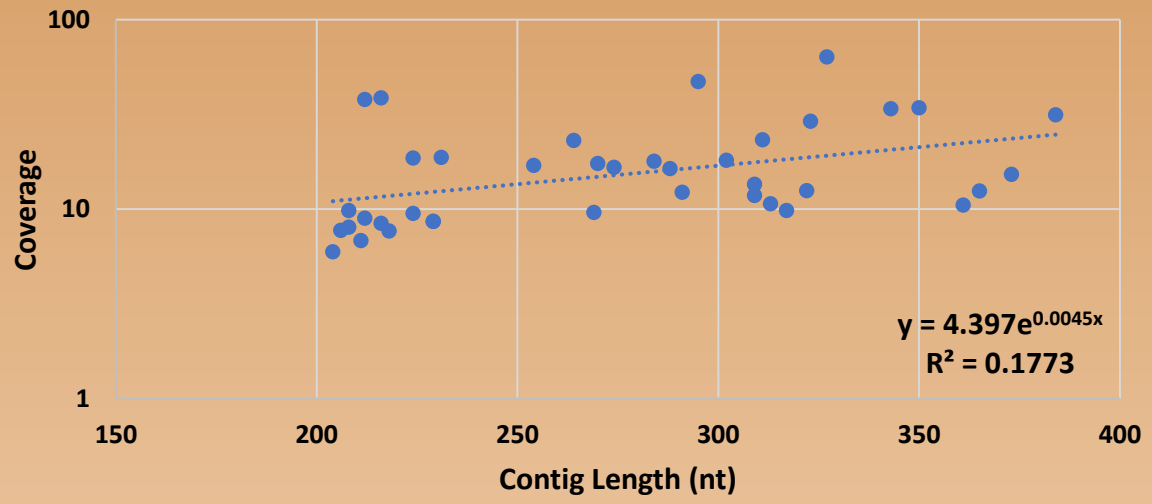
Coverage vs Contig Length (Novel lncRNA) Without "Outliers"



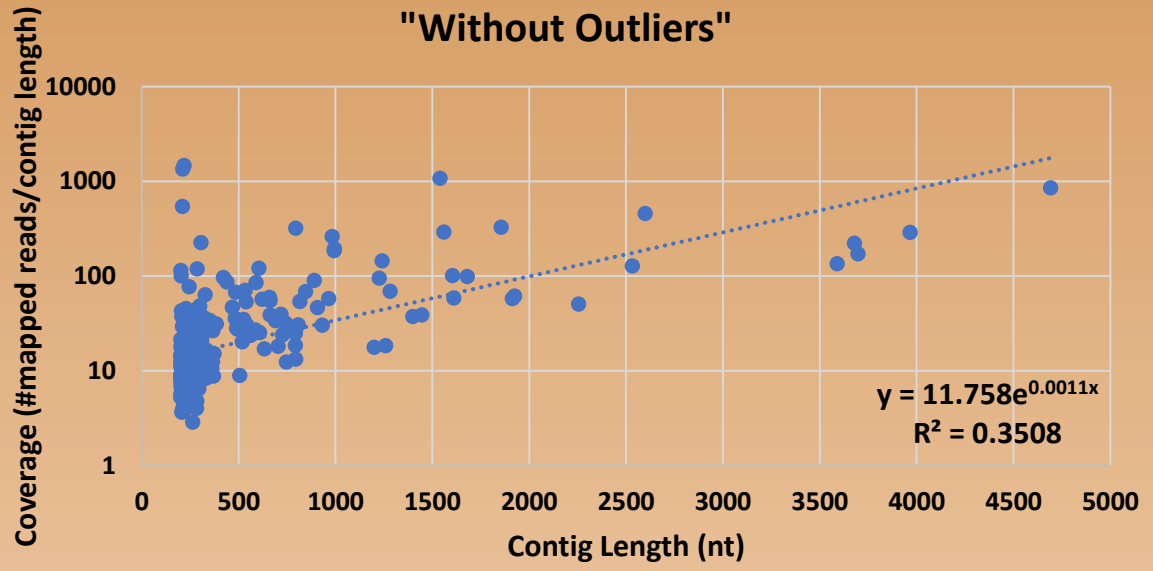
Coverage vs Contig Length (Known lncRNA) Without "Outliers"



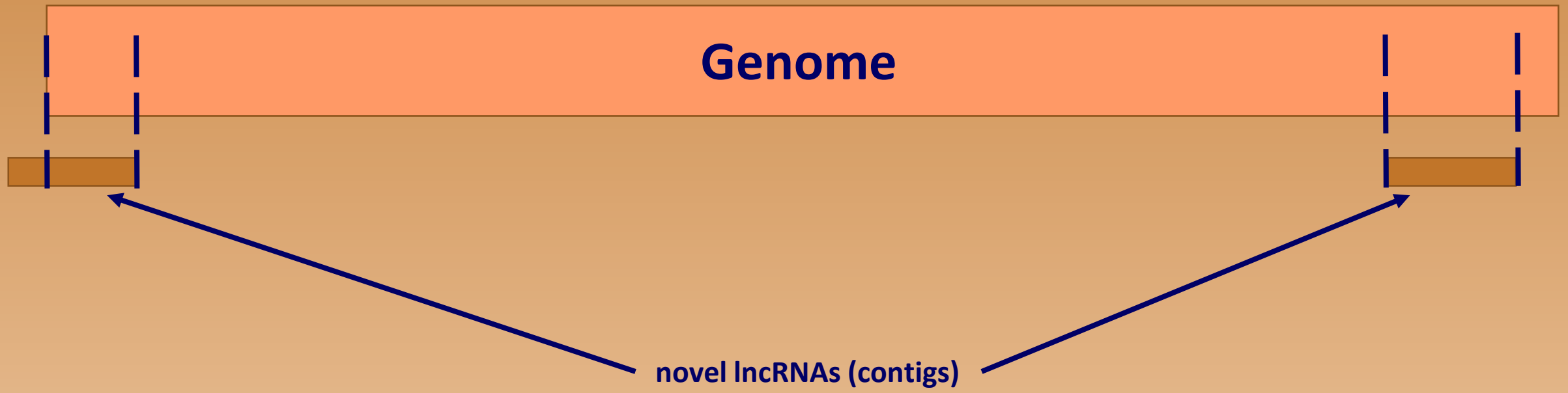
Coverage vs Contig Length (40th Percentile lncRNA) Without "Outliers"



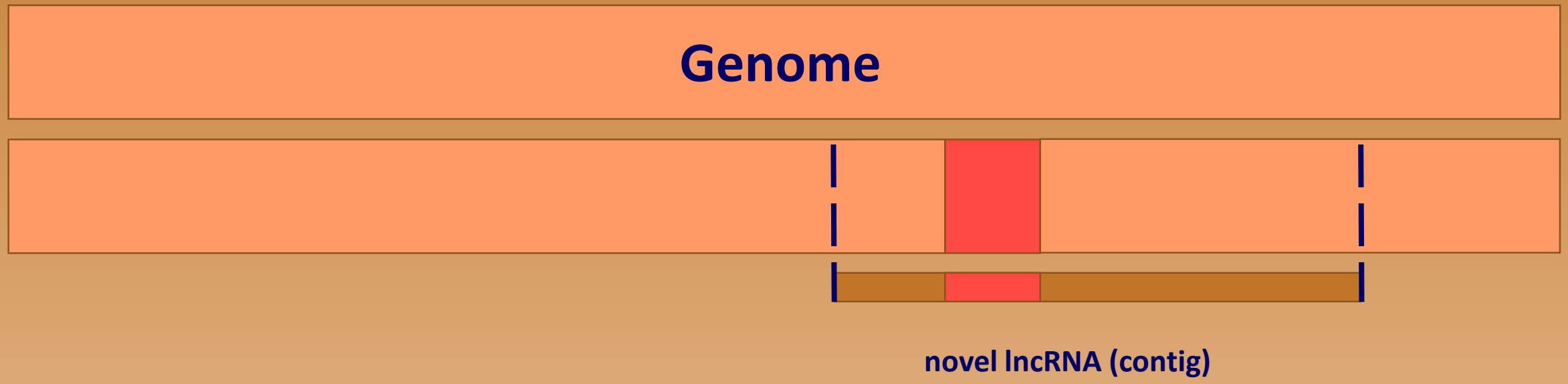
Coverage vs Contig Length (Pooled lncRNA) "Without Outliers"



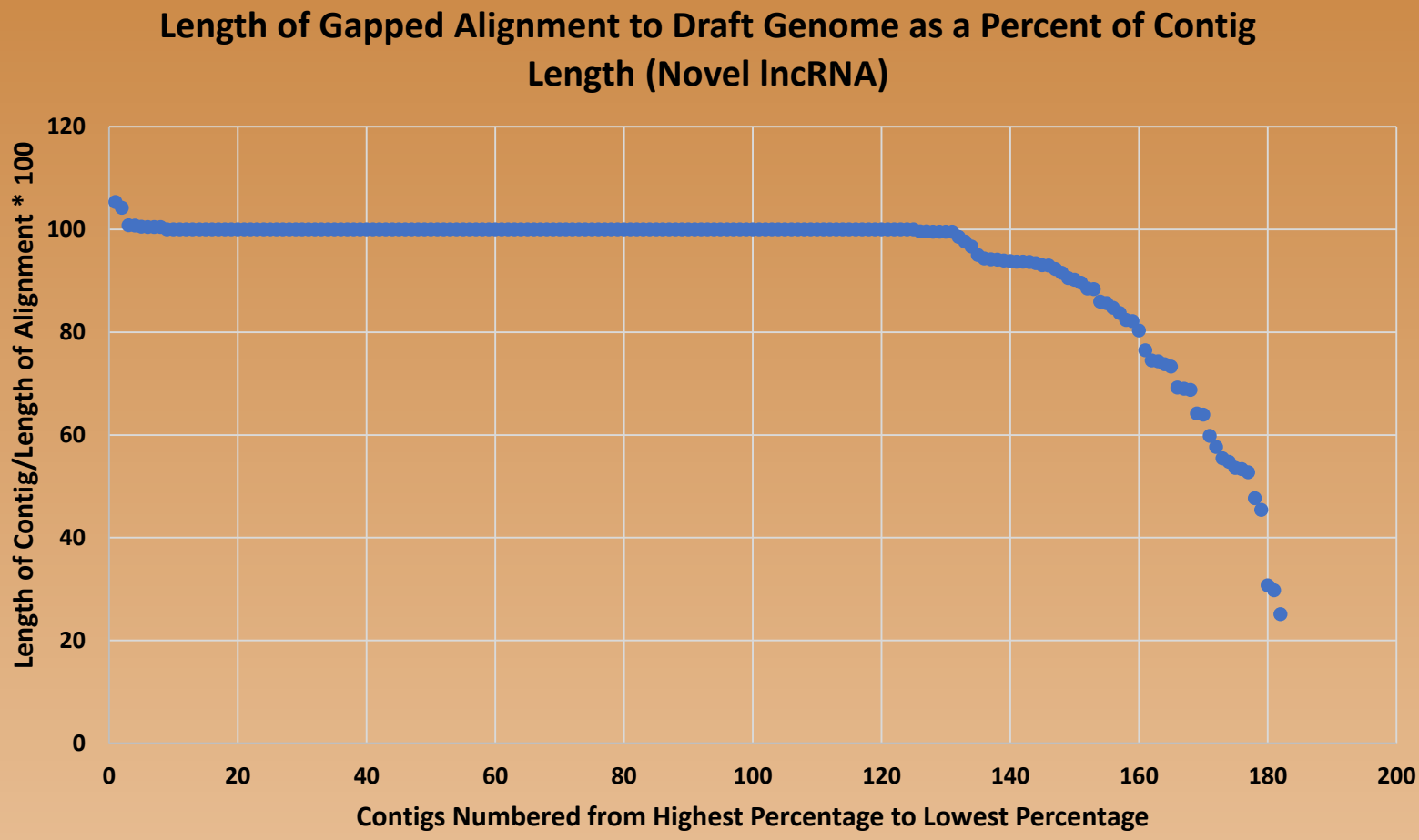
The novel lncRNA map to the genome



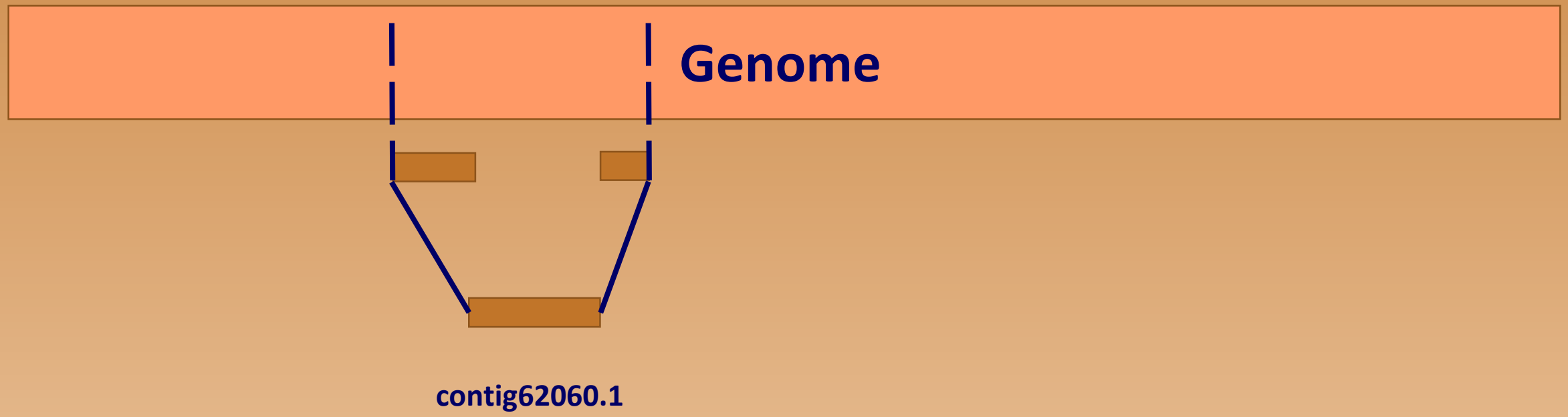
The match is not necessarily perfect



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

	Measure				
Contig	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

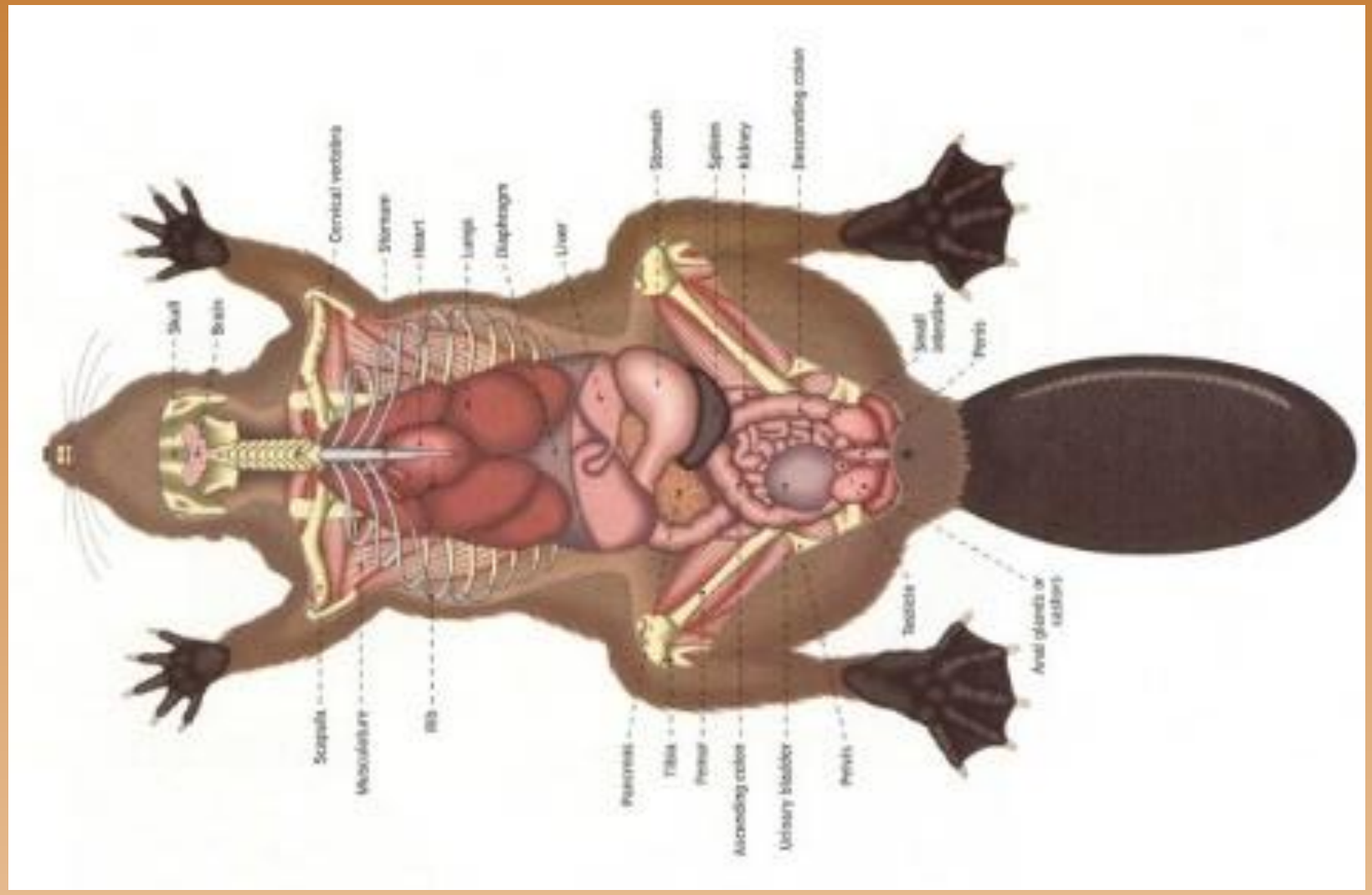


Image Credit: crappie.com, Pinterest. *Inside look at the Beaver.....*

Acknowledgments

Oregon Zoo

Mitch Finnegan

Tim Storms

Amy Cutting

David Shepherdson

CENTER FOR GENOME RESEARCH & BIOCOMPUTING

THANK YOU