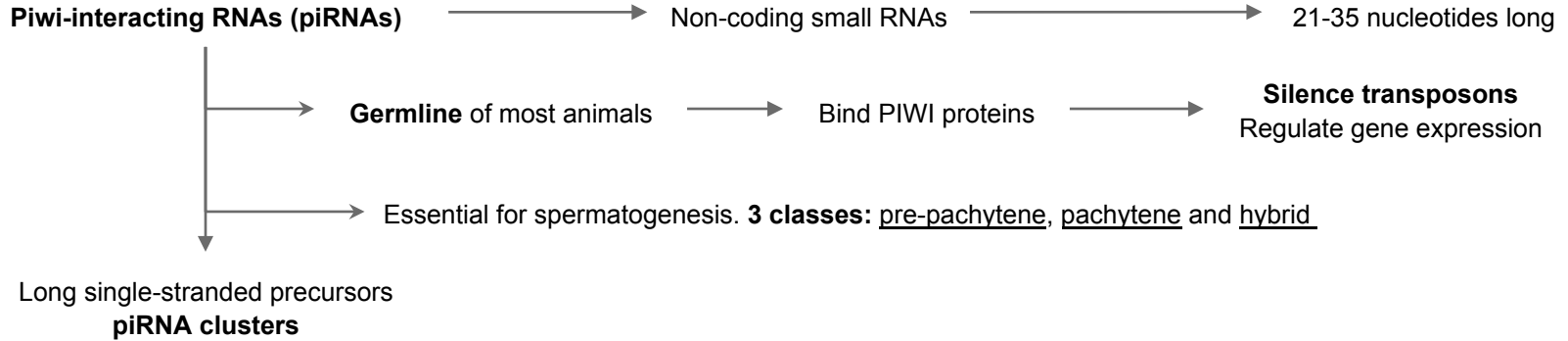


# Analysis of the variation in piRNA expression in the male germline of three *Mus* species

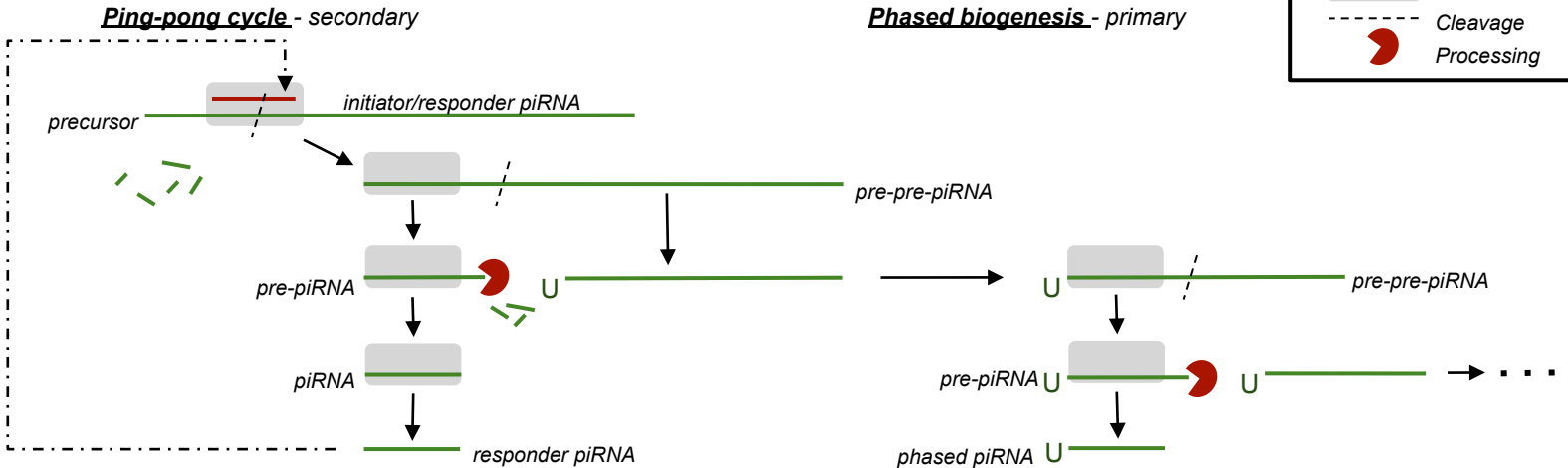
**Adrià Mitjavila Ventura**  
MSc in Bioinformatics and Biostatistics  
Universitat Oberta de Catalunya

January 3, 2022

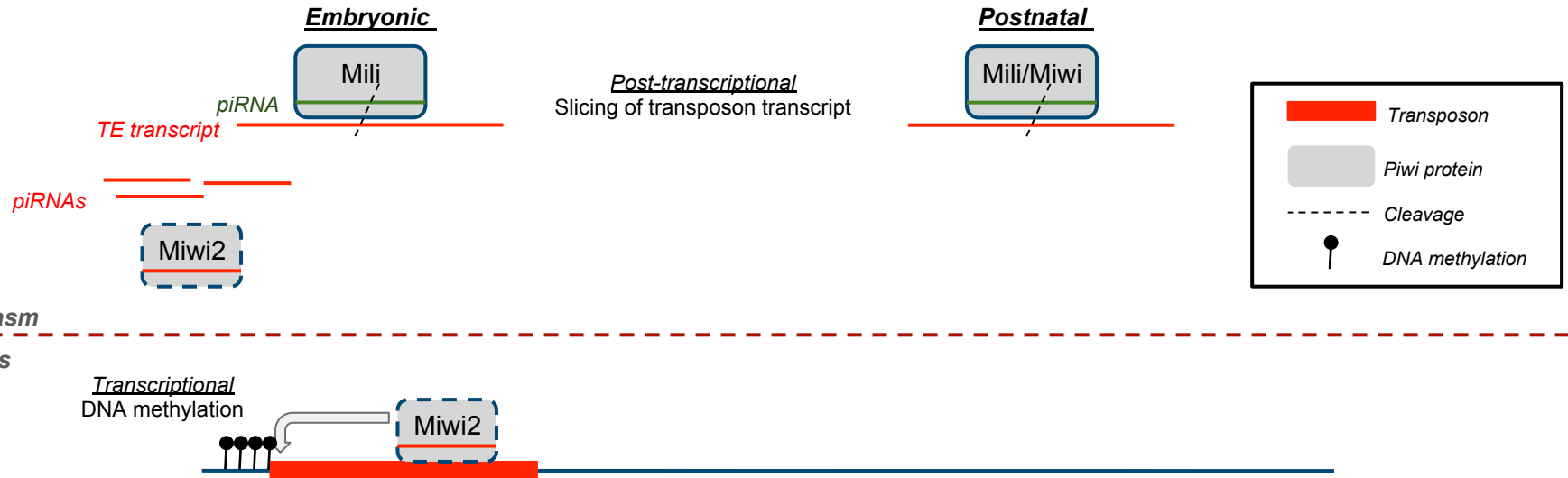
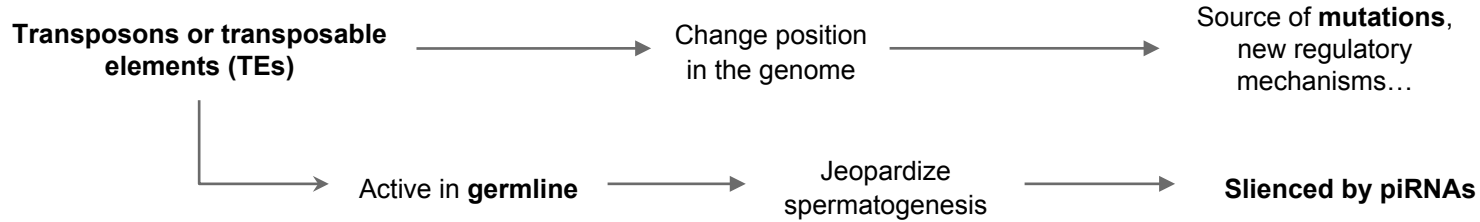
# Introduction - piRNAs



mitochondria outer membrane



# Introduction - Transposons and piRNAs



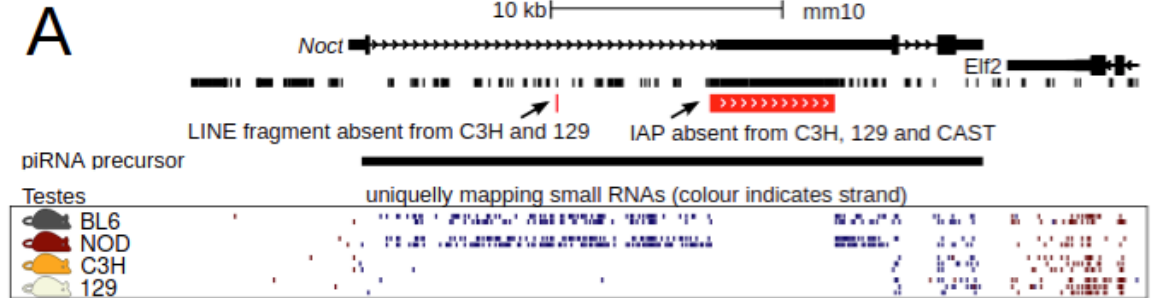
# Question - What explains piRNA variation?

- ❗ piRNAs are essential, but vary a lot between species.
- ? The signal that selects a locus for piRNA production is not known!
- ❗ Signal must be within the cluster

Active piRNA cluster → Ectopic location → Still active → Genomic location is not important

## Hypothesis

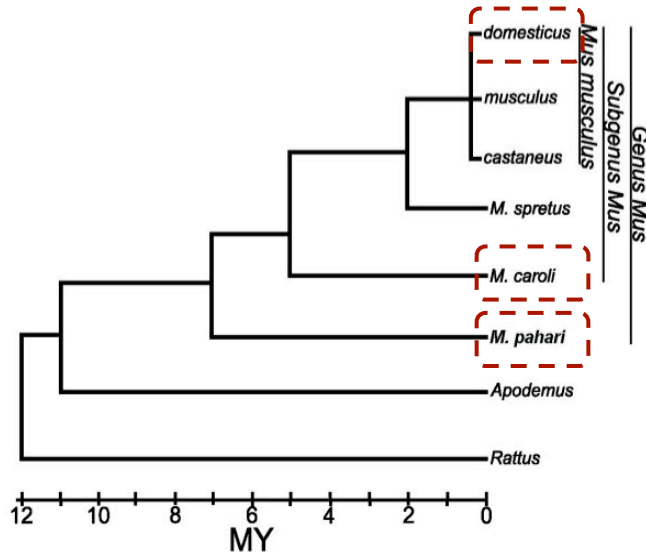
TEs interact with piRNA biogenesis machinery and induce piRNA production



*piRNAs from Noct in mouse strains depending on IAP insertion. Tanya Vavouri, personal communication*

# Methods - Samples

Species	Replicates	Age (weeks)	Tissue	Genome assembly	Distance to <i>M. musculus</i>
<i>Mus musculus</i>	2	14	Whole testis	GRCm39	0 MY
<i>Mus caroli</i>	2	9-10	Whole testis	CAROLI_EIJ_v1.1	3-5 MY
<i>Mus pahari</i>	2	8	Whole testis	PAHARI_EIJ_v1.1	6-7 MY



Wikimedia commons // George Shuklin



K. Tsuchiya // <https://alchetron.com/Ryukyuu-mouse>



Adapted from: Wikimedia commons // Priya Tamma

Phylogenetic tree of some Murinae species.  
Laukaitis et al., 2008. doi: 10.1186/1471-2148-8-46

# Methods - piRNA cluster annotation

## Previously defined (*Yu et al* clusters)

*Yu et al., 2021.*

doi: 10.1038/s41467-020-20345-3

*Mus musculus*  
Postnatal testis  
Several developmental stages  
Seq: RNA, small RNA, ChIP(s), CAGE, PAS...

467 transcripts

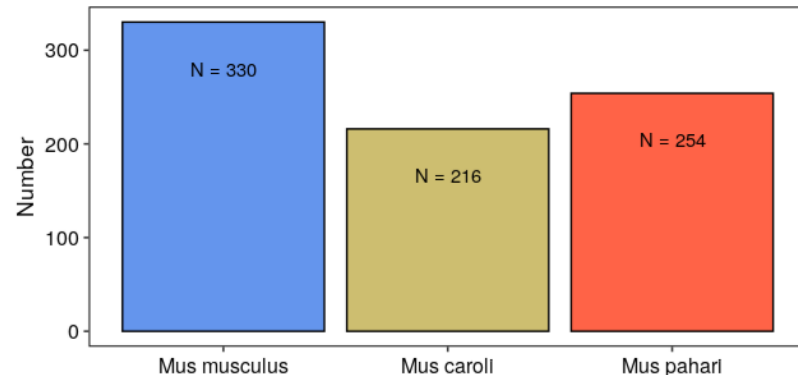
**215 clusters**  
**Gold standard**

## Predicted *de novo* with *proTRAC* (*de novo* clusters)

*Mus musculus, Mus caroli* or *Mus pahari*  
Adult testis  
small RNA-seq

Predict clusters with  
*proTRAC*

***de novo* clusters**



Number of *de novo* clusters predicted in each species

# Challenge - Orthologs and differential expression

## Assumption

Equal length

cond1   
cond2 

## Reality

Different species → different lengths


## Solution

Conserved blocks from  
ENSEMBL Compara  
Murinae multiple alignment

Original (GRCm39)

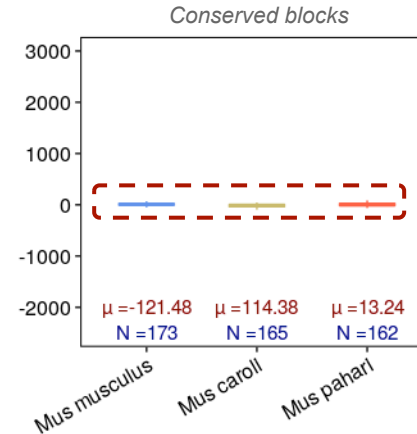
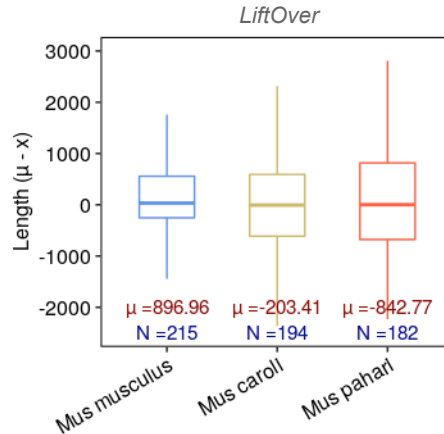
chr10 105 123 1 region1

ENSEMBL  
Compara Perl API  
→  
Bash + R

chr10 105 123 1 region1  
chr10 86 102 1 region1  
chr9 157 172 -1 region1

Orthologs

GRCm39 **AATCGT-ATCGT-AA-T-GGCT**  
CAROLI **AATC--CGTCG-GGTG---CCT**  
PAHARI **AAT---T-TGG-GAT-TC-GCT**

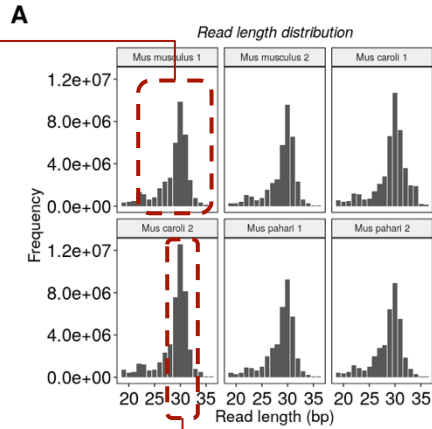


Differences between mean length across the three species and length in each species.

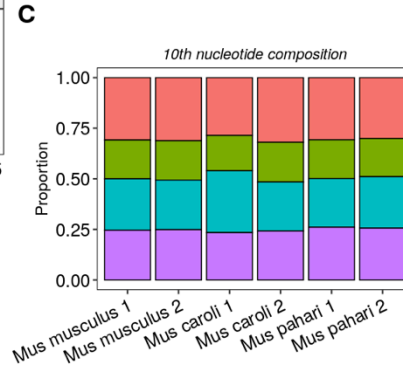
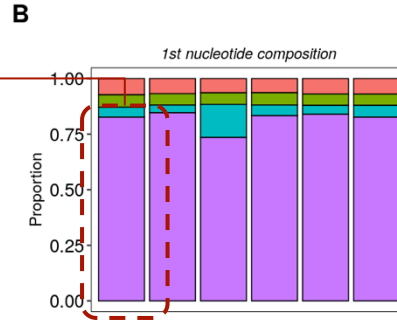
# Quality control - Our data is enriched in piRNAs

1st nucleotide is U =  
phased biogenesis

Most reads between  
21-35 nucleotides (nuc)



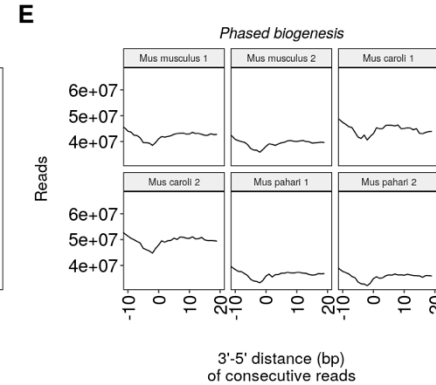
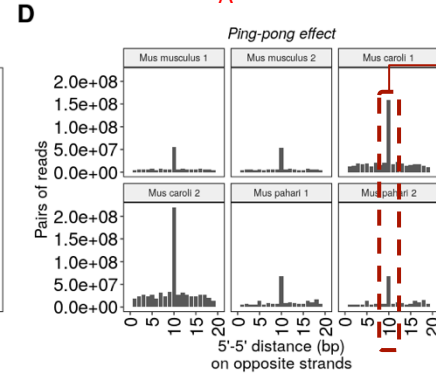
Peak at 30 nuc. =  
piRNAs bound to MIWI



■ A ■ C ■ G ■ U



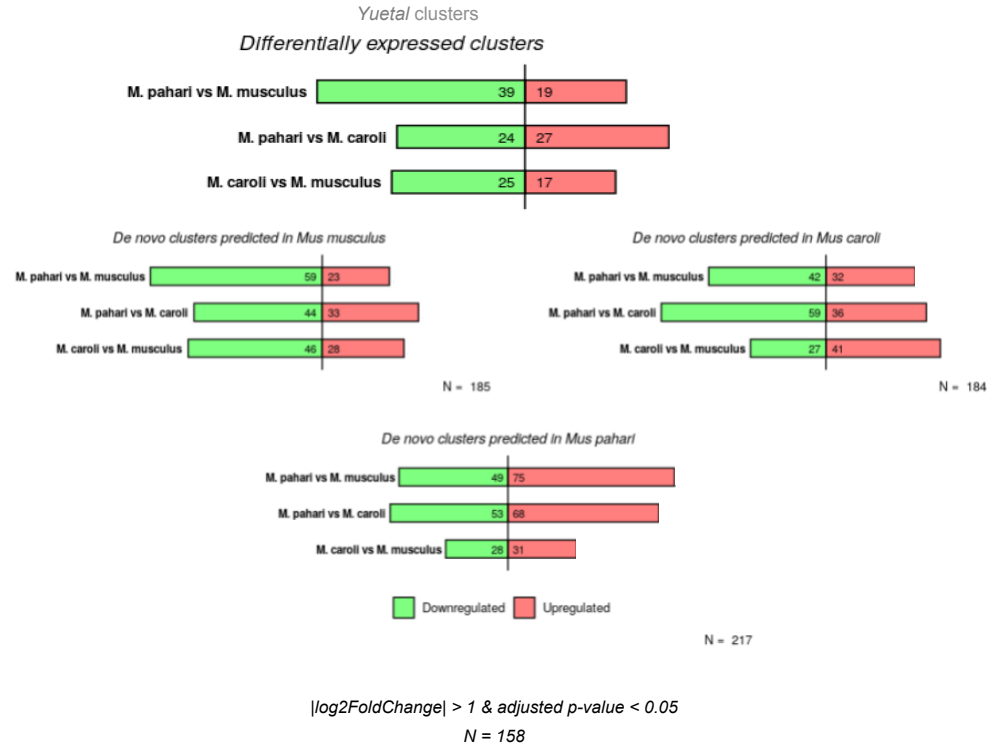
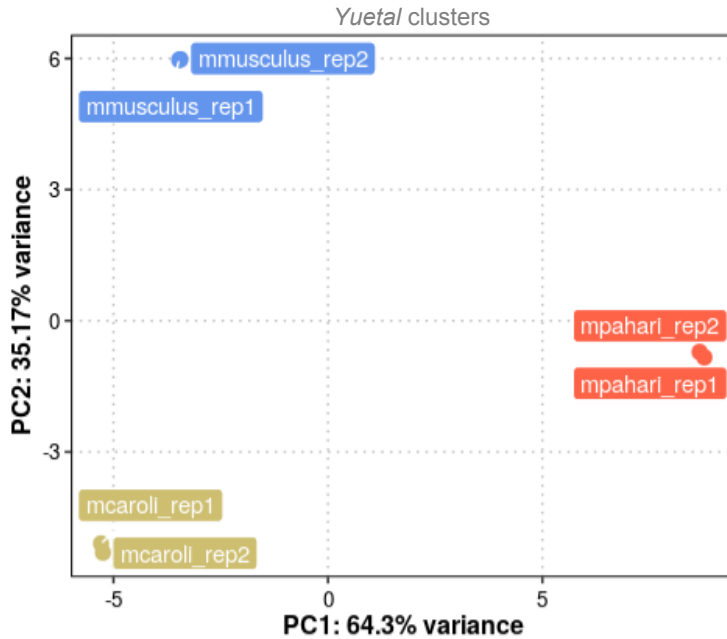
Peak at 10 nuc. = ping-  
pong amplification





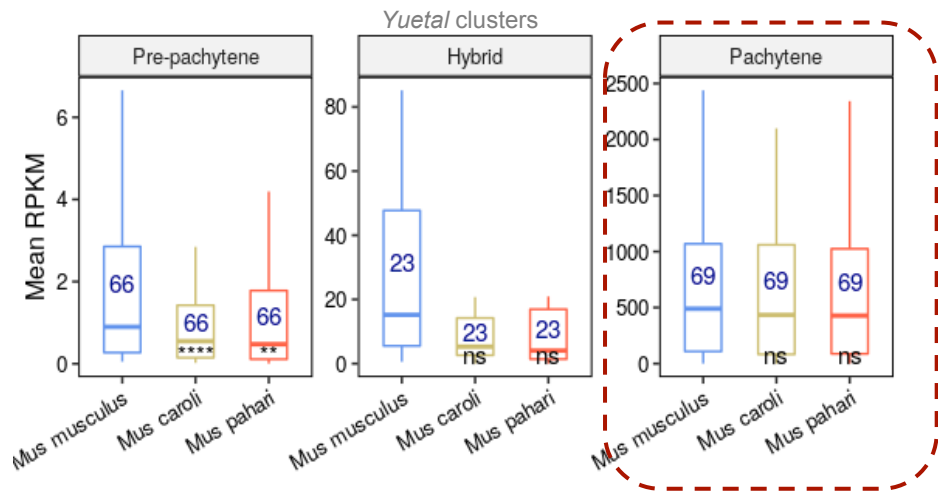
# Results - piRNAs expression varies greatly across species

How much piRNA expression varies across the three species?



# Results - Pachytene clusters are more conserved (in expression)

Are piRNAs from distinct classes expressed differently?



Yuetai clusters  
Pre-pachytene piRNA clusters

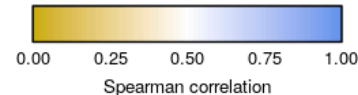
Mus musculus 1						1
Mus musculus 2					1	0.99
Mus caroli 1				1	0.75	0.74
Mus caroli 2				1	0.99	0.76
Mus pahari 1			1	0.74	0.74	0.71
Mus pahari 2	1	1	0.74	0.74	0.7	0.67

Hybrid piRNA clusters

Mus musculus 1						1
Mus musculus 2					1	1
Mus caroli 1				1	0.75	0.76
Mus caroli 2				1	1	0.76
Mus pahari 1		1	0.64	0.63	0.45	0.46
Mus pahari 2	1	1	0.63	0.63	0.46	0.47

Pachytene piRNA clusters

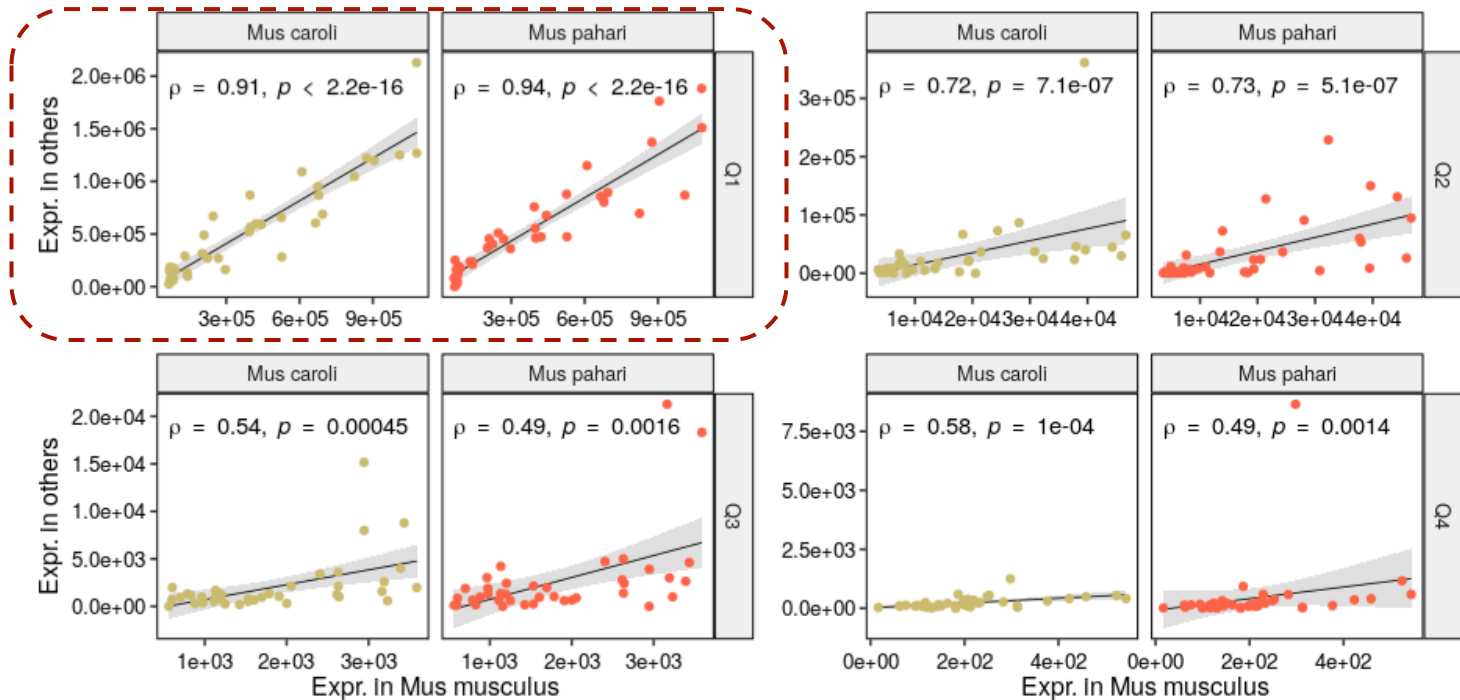
Mus musculus 1						1
Mus musculus 2					1	1
Mus caroli 1				1	0.95	0.95
Mus caroli 2				1	1	0.95
Mus pahari 1		1	0.91	0.91	0.95	0.95
Mus pahari 2	1	1	0.91	0.91	0.95	0.95



# Results - Expression correlates with conservation (of expression)

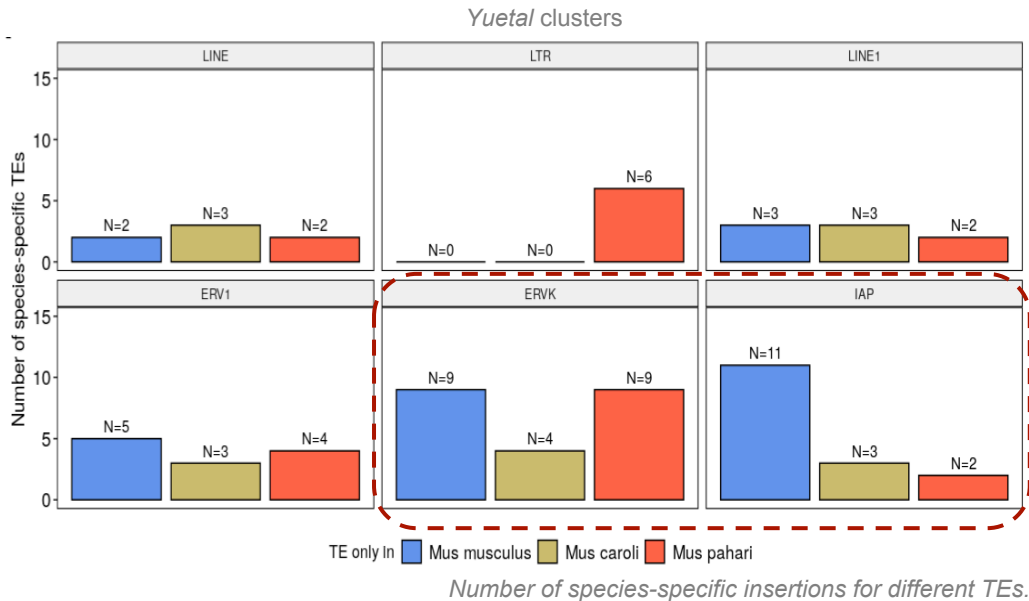
*Does expression correlates with conservation?*

Yu et al clusters



# Results - TEs do not explain global changes in piRNA expression

*Do TEs explain the changes in piRNA expression across species?*



Very few compared to DE clusters

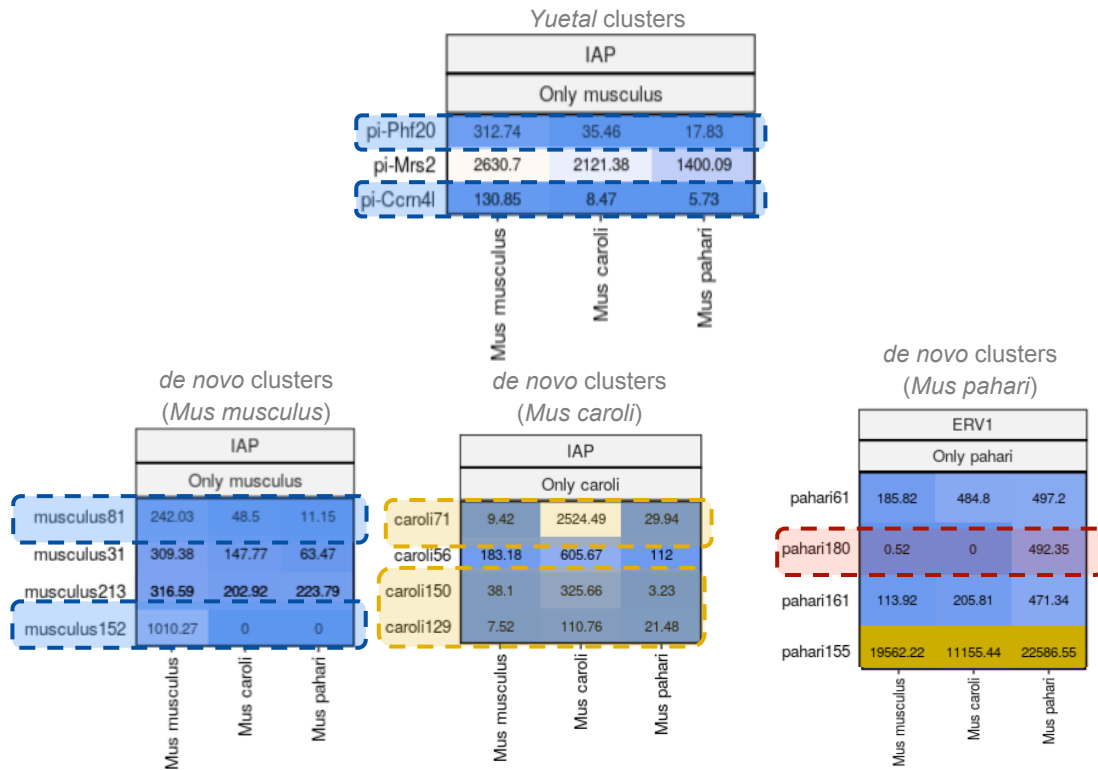
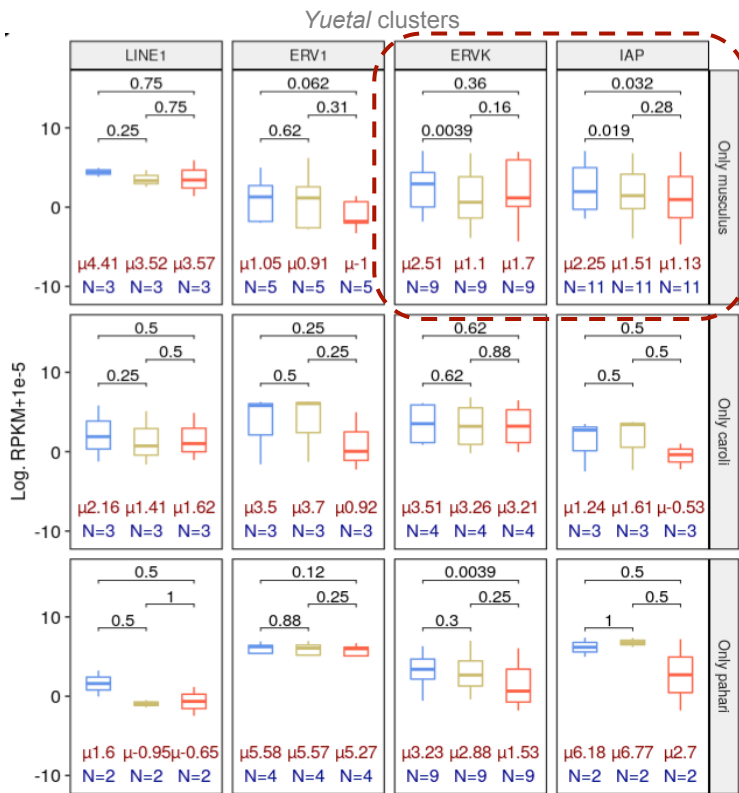
		<i>Is cluster differentially expressed?</i>	
		No	Yes
<i>Is TE only in one species?</i>	No	x	x
	Yes	x	x

*Fisher's tests*

<b><i>Yuetal clusters</i></b>
<i>No significant associations</i>
<b><i>de novo clusters from Mus musculus</i></b>
<i>ERV1 in Mus musculus (<math>p &lt; 0.05</math>)</i> <i>LINE1 in Mus pahari (<math>p &lt; 0.1</math>)</i> <i>ERV1 in Mus pahari (<math>p &lt; 0.05</math>)</i>
<b><i>de novo clusters from Mus caroli</i></b>
<i>ERVK in Mus musculus (<math>p &lt; 0.1</math>)</i>
<b><i>de novo clusters from Mus pahari</i></b>
<i>LINE1 in Mus pahari (<math>p &lt; 0.05</math>)</i>

# Results - Few significant changes in clusters with TEs

How much does the expression change in clusters with TEs?



Expression in Mus musculus Mus caroli Mus pahari

Manual inspection of some clusters  
High expression in the species with TE  
Very low expression in the others

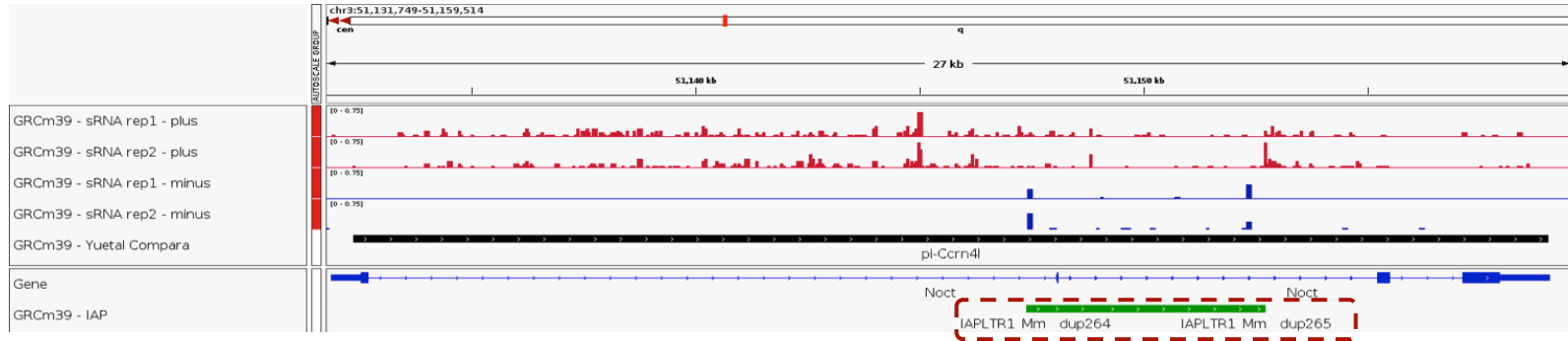
# Results - 2 clusters with IAP

**Noct**: genic cluster

Previously associated with IAP

Sense intronic IAP in *Mus musculus*

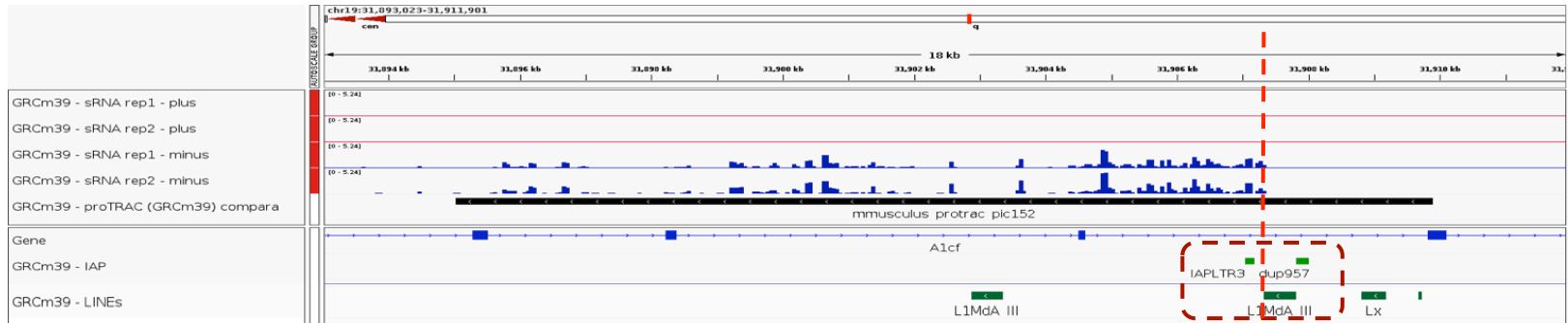
piRNAs in introns → **interaction of IAP with splicing machinery?**



**musculus152**: intergenic cluster

Antisense IAP in *Mus musculus*

piRNAs downstream of IAP → **target for initiator piRNA?**



# Summary

---



## CONCLUSIONS

- Our sRNA data was highly enriched in piRNAs.
- More conserved clusters (in terms of expression) produced more piRNAs.
- Transposons **alone** do not explain **global** changes in piRNA expression.
- Some examples are consistent with the proposed model.



## QUESTIONS

- Which is the mechanism linking IAP/TEs to piRNA production?
- Are there differences between genic and intergenic clusters?



## FUTURE RESEARCH

- Include mouse strains.

# Thank you!

---

## Special acknowledgements

### **IJC**

#### **Vavouri lab**

Tanya Vavouri

Me

### **HR**

Carolina Carmona

### **Tutor**

Elisabetta Mereu

### **IGTP**

#### **Genomics Unit**



*Wikimedia commons // George Shuklin*

---

## References

Laukaitis C, Heger A, Blakley T, Munclinger P, Ponting C, Karn R. Rapid bursts of androgen-binding protein (Abp) gene duplication occurred independently in diverse mammals. *BMC Evolutionary Biology*. 2008;8. 46. doi: 10.1186/1471-2148-8-46.

Yu T, Fan K, Ozata DM, Zhang G, Fu Y, Theurkauf WE, Zamore PD, Weng Z. Long first exons and epigenetic marks distinguish conserved pachytene piRNA clusters from other mammalian genes. *Nat Commun*. 2021 Jan 4;12(1):73. doi: 10.1038/s41467-020-20345-3. PMID: 33397987; PMCID: PMC7782496.