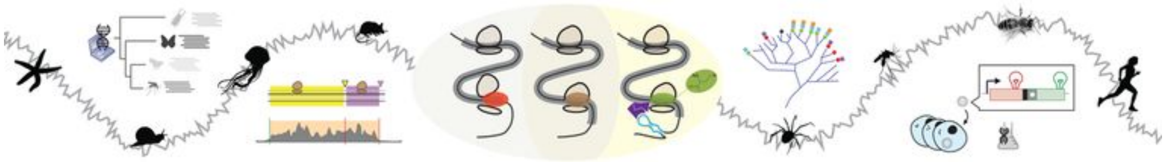
MARIOTTI LAB @ UB
COMPARATIVE GENOMICS

UNIVERSITAT DE BARCELONA

Department of Genetics, Microbiology and Statistics

Decoding the determinants of stop codon readthrough in insects

Nadezhda (Nadya) Makarova
PhD student
Mariotti Lab / Comparative Genomics and Recoding

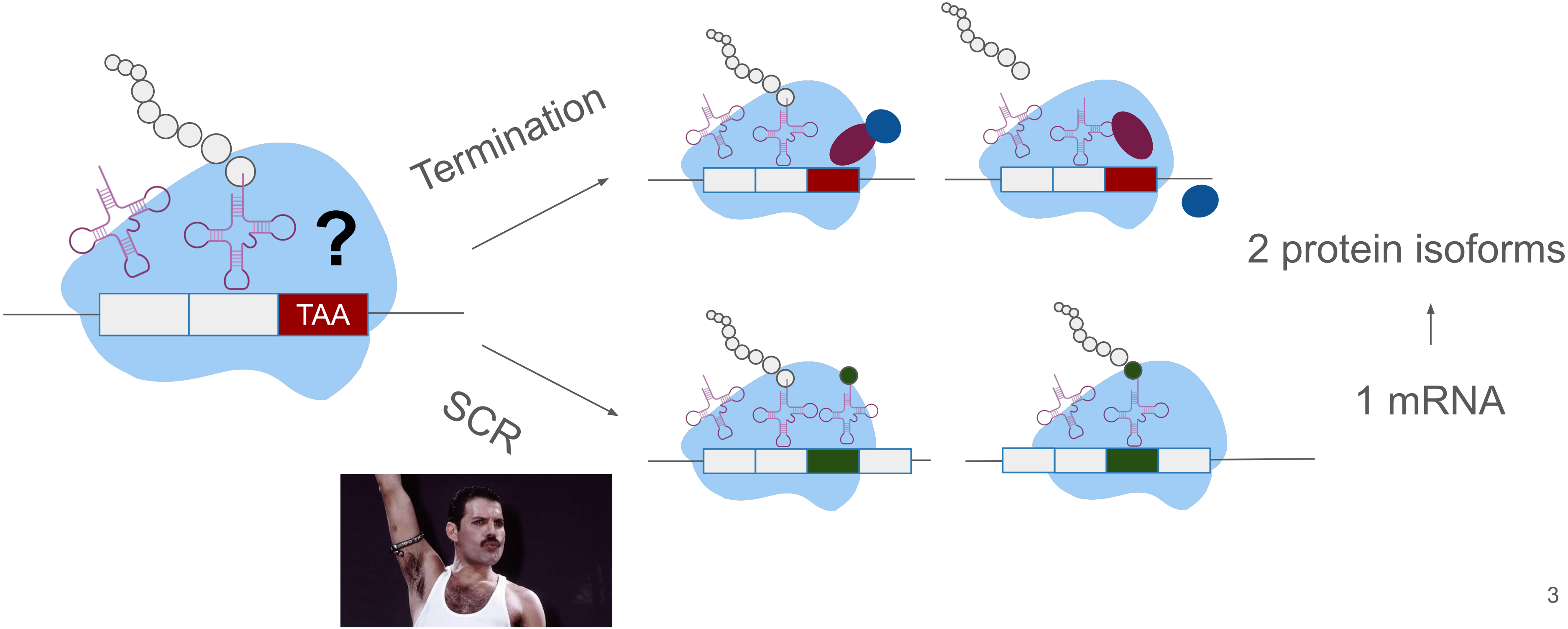


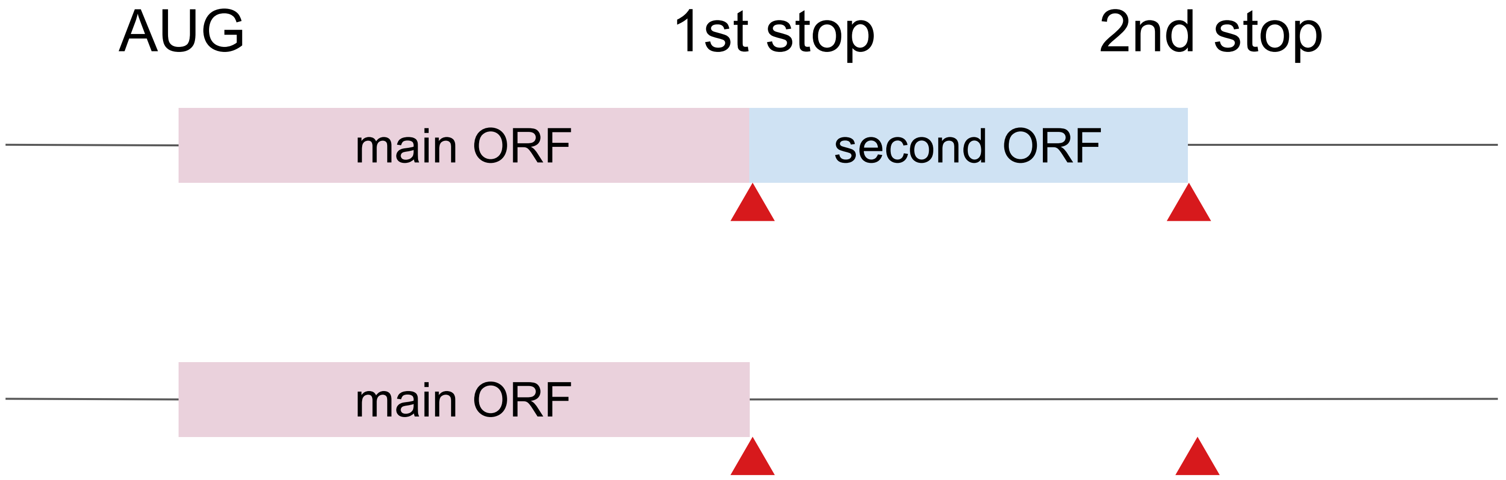
Benasque, 31.07.2024

How to characterise not annotated but functional RNA structures?

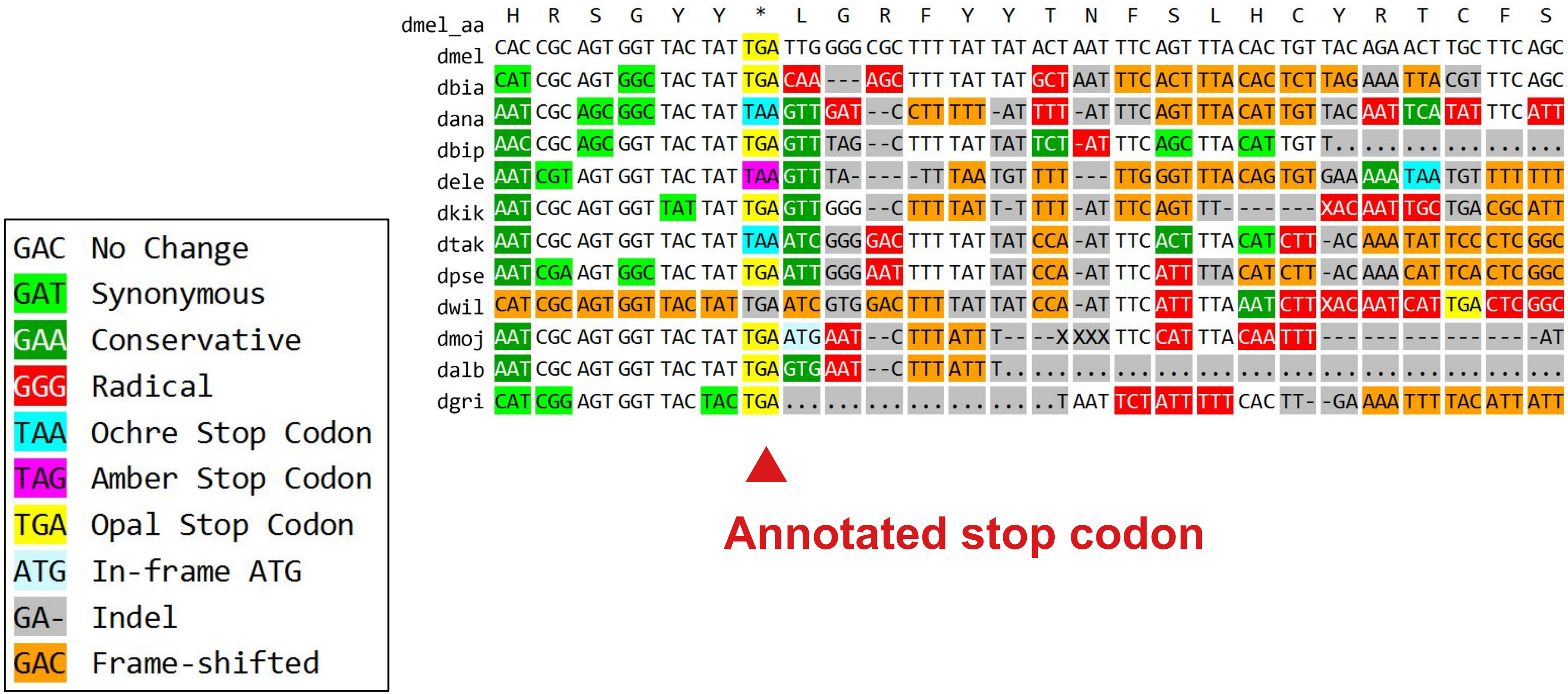
Stop Codon Readthrough (SCR)

A form of recoding= programmed exceptions to the genetic code




AUG
1st stop
2nd stop
main ORF
second ORF
main ORF

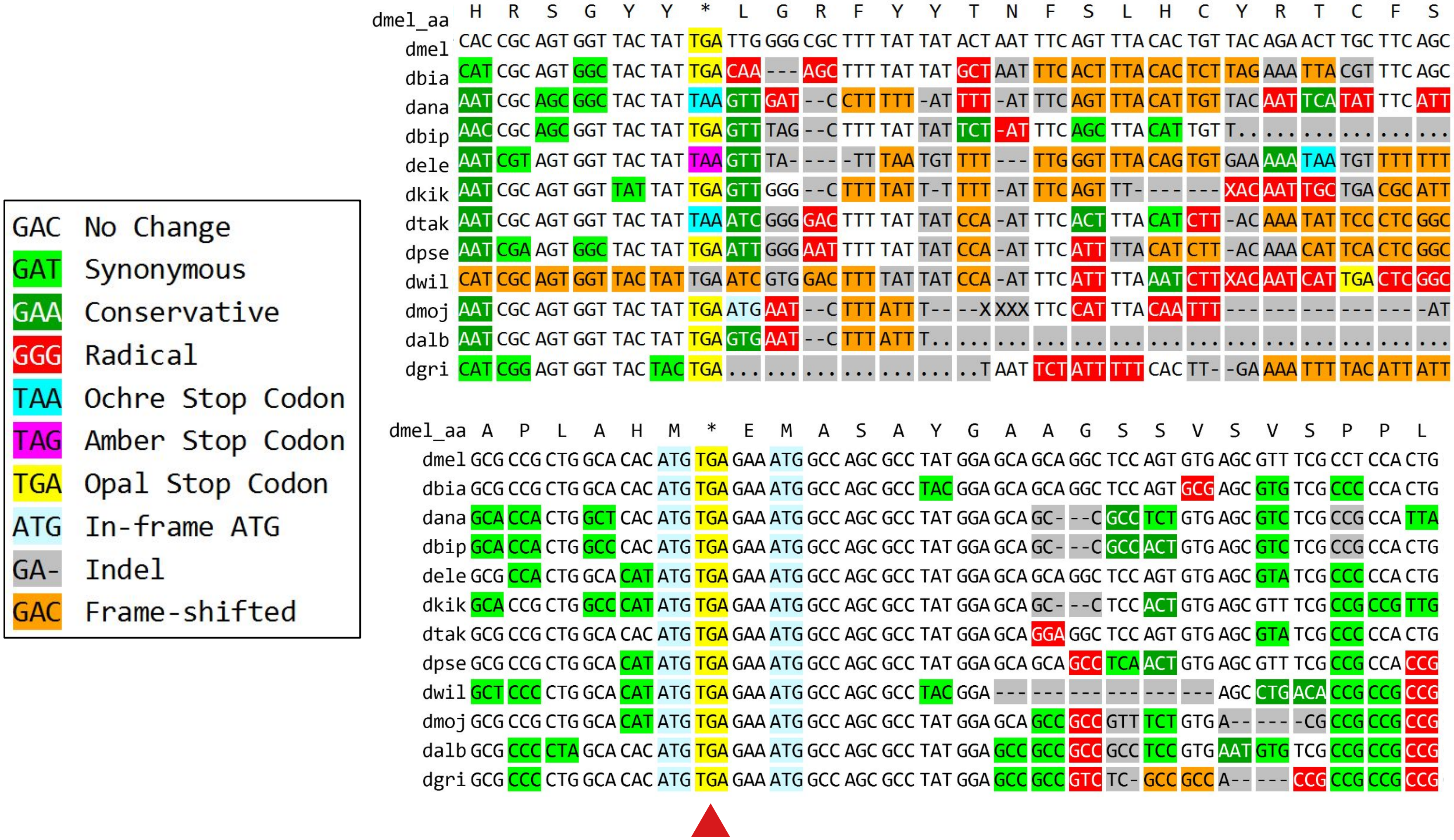
Gene 2



dmel_aa	H	R	S	G	Y	Y	*	L	G	R	F	Y	Y	T	N	F	S	L	H	C	Y	R	T	C	F	S
dmel	CAC	CGC	AGT	GGT	TAC	TAT	TGA	TTG	GGG	CGC	TTT	TAT	TAT	ACT	AAT	TTC	AGT	TTA	CAC	TGT	TAC	AGA	ACT	TGC	TTC	AGC
dbia	CAT	CGC	AGT	GGC	TAC	TAT	TGA	CAA	---	AGC	TTT	TAT	TAT	GCT	AAT	TTC	ACT	TTA	CAC	TCT	TAG	AAA	TTA	CGT	TTC	AGC
dana	AAT	CGC	AGC	GGC	TAC	TAT	TAA	GTT	GAT	--C	CTT	TTT	-AT	TTT	-AT	TTC	AGT	TTA	CAT	TGT	TAC	AAT	TCA	TAT	TTC	ATT
dbip	AAC	CGC	AGC	GGT	TAC	TAT	TGA	GTT	TAG	--C	TTT	TAT	TAT	TCT	-AT	TTC	AGC	TTA	CAT	TGT	T..	...	...	...	...	...
dele	AAT	CGT	AGT	GGT	TAC	TAT	TAA	GTT	TA-	---	-TT	TAA	TGT	TTT	---	TTG	GGT	TTA	CAG	TGT	GAA	AAA	TAA	TGT	TTT	TTT
dkik	AAT	CGC	AGT	GGT	TAT	TAT	TGA	GTT	GGG	--C	TTT	TAT	T-T	TTT	-AT	TTC	AGT	TT-	---	---	XAC	AAT	TGC	TGA	CGC	ATT
dtak	AAT	CGC	AGT	GGT	TAC	TAT	TAA	ATC	GGG	GAC	TTT	TAT	TAT	CCA	-AT	TTC	ACT	TTA	CAT	CTT	-AC	AAA	TAT	TCC	CTC	GGC
dpse	AAT	CGA	AGT	GGC	TAC	TAT	TGA	ATT	GGG	AAT	TTT	TAT	TAT	CCA	-AT	TTC	ATT	TTA	CAT	CTT	-AC	AAA	CAT	TCA	CTC	GGC
dwil	CAT	CGC	AGT	GGT	TAC	TAT	TGA	ATC	GTG	GAC	TTT	TAT	TAT	CCA	-AT	TTC	ATT	TTA	AAT	CTT	XAC	AAT	CAT	TGA	CTC	GGC
dmoj	AAT	CGC	AGT	GGT	TAC	TAT	TGA	ATG	AAT	--C	TTT	ATT	T--	--X	XXX	TTC	CAT	TTA	CAA	TTT	---	---	---	---	---	-AT
dalb	AAT	CGC	AGT	GGT	TAC	TAT	TGA	GTG	AAT	--C	TTT	ATT	T..	...	...	...	...	...	...	...	...	...	...	...	...	...
dgri	CAT	CGG	AGT	GGT	TAC	TAC	TGA	...	...	...	...	...	...	..T	AAT	TCT	ATT	TTT	CAC	TT-	-GA	AAA	TTT	TAC	ATT	ATT

GAC	No Change
GAT	Synonymous
GAA	Conservative
GGG	Radical
TAA	Ochre Stop Codon
TAG	Amber Stop Codon
TGA	Opal Stop Codon
ATG	In-frame ATG
GA-	Indel
GAC	Frame-shifted

Annotated stop codon



GAC	No Change
GAT	Synonymous
GAA	Conservative
GGG	Radical
TAA	Ochre Stop Codon
TAG	Amber Stop Codon
TGA	Opal Stop Codon
ATG	In-frame ATG
GA-	Indel
GAC	Frame-shifted

dmel_aa	H	R	S	G	Y	Y	*	L	G	R	F	Y	Y	T	N	F	S	L	H	C	Y	R	T	C	F	S
dmel	CAC	CGC	AGT	GGT	TAC	TAT	TGA	TTG	GGG	CGC	TTT	TAT	TAT	ACT	AAT	TTC	AGT	TTA	CAC	TGT	TAC	AGA	ACT	TGC	TTC	AGC
dbia	CAT	CGC	AGT	GGC	TAC	TAT	TGA	CAA	---	AGC	TTT	TAT	TAT	GCT	AAT	TTC	ACT	TTA	CAC	TCT	TAG	AAA	TTA	CGT	TTC	AGC
dana	AAT	CGC	AGC	GGC	TAC	TAT	TAA	GTT	GAT	--C	CTT	TTT	-AT	TTT	-AT	TTC	AGT	TTA	CAT	TGT	TAC	AAT	TCA	TAT	TTC	ATT
dbip	AAC	CGC	AGC	GGT	TAC	TAT	TGA	GTT	TAG	--C	TTT	TAT	TAT	TCT	-AT	TTC	AGC	TTA	CAT	TGT	T..	...	...	...	...	...
dele	AAT	CGT	AGT	GGT	TAC	TAT	TAA	GTT	TA-	---	-TT	TAA	TGT	TTT	---	TTG	GGT	TTA	CAG	TGT	GAA	AAA	TAA	TGT	TTT	TTT
dkik	AAT	CGC	AGT	GGT	TAT	TAT	TGA	GTT	GGG	--C	TTT	TAT	T-T	TTT	-AT	TTC	AGT	TT-	---	---	XAC	AAT	TGC	TGA	CGC	ATT
dtak	AAT	CGC	AGT	GGT	TAC	TAT	TAA	ATC	GGG	GAC	TTT	TAT	TAT	CCA	-AT	TTC	ACT	TTA	CAT	CTT	-AC	AAA	TAT	TCC	CTC	GGC
dpse	AAT	CGA	AGT	GGC	TAC	TAT	TGA	ATT	GGG	AAT	TTT	TAT	TAT	CCA	-AT	TTC	ATT	TTA	CAT	CTT	-AC	AAA	CAT	TCA	CTC	GGC
dwil	CAT	CGC	AGT	GGT	TAC	TAT	TGA	ATC	GTG	GAC	TTT	TAT	TAT	CCA	-AT	TTC	ATT	TTA	AAT	CTT	XAC	AAT	CAT	TGA	CTC	GGC
dmoj	AAT	CGC	AGT	GGT	TAC	TAT	TGA	ATG	AAT	--C	TTT	ATT	T--	--X	XXX	TTC	CAT	TTA	CAA	TTT	---	---	---	---	---	-AT
dalb	AAT	CGC	AGT	GGT	TAC	TAT	TGA	GTG	AAT	--C	TTT	ATT	T..	...	...	...	...	...	...	...	...	...	...	...	...	...
dgri	CAT	CGG	AGT	GGT	TAC	TAC	TGA	...	...	...	...	...	...	..T	AAT	TCT	ATT	TTT	CAC	TT-	-GA	AAA	TTT	TAC	ATT	ATT

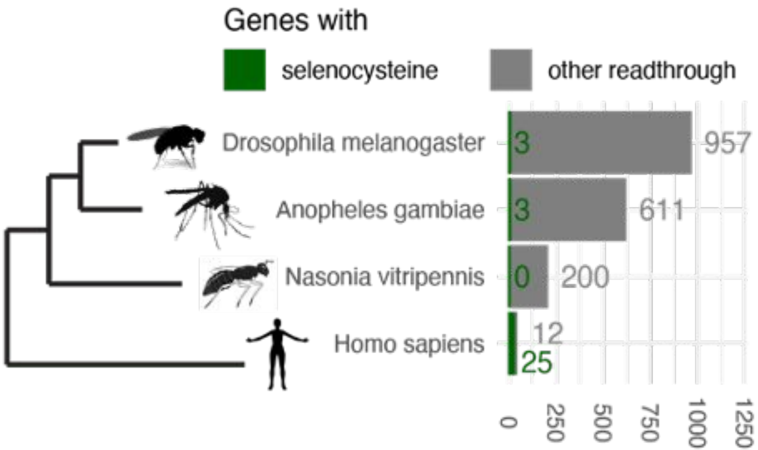
Gene 2

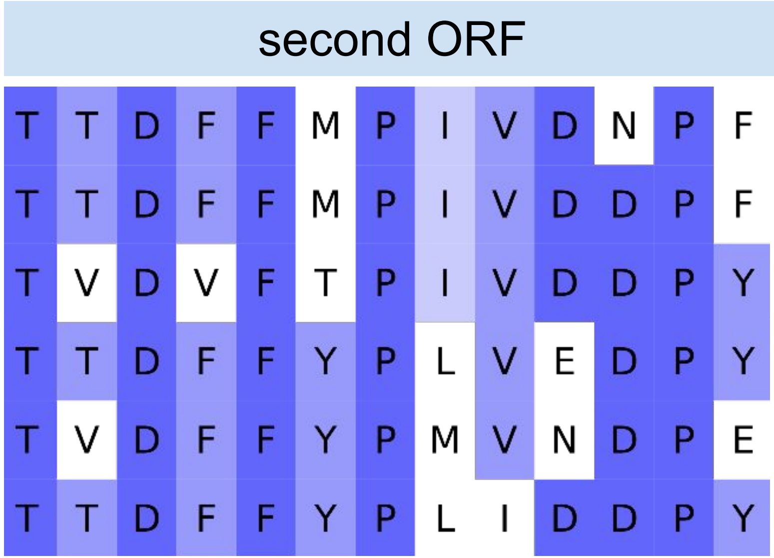
dmel_aa	A	P	L	A	H	M	*	E	M	A	S	A	Y	G	A	A	G	S	S	V	S	V	S	P	P	L
dmel	GCG	CCG	CTG	GCA	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GCA	GGC	TCC	AGT	GTG	AGC	GTT	TCG	CCT	CCA	CTG
dbia	GCG	CCG	CTG	GCA	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAC	GGA	GCA	GCA	GGC	TCC	AGT	GCG	AGC	GTG	TCG	CCC	CCA	CTG
dana	GCA	CCA	CTG	GCT	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GC-	--C	GCC	TCT	GTG	AGC	GTC	TCG	CCG	CCA	TTA
dbip	GCA	CCA	CTG	GCC	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GC-	--C	GCC	ACT	GTG	AGC	GTC	TCG	CCG	CCA	CTG
dele	GCG	CCA	CTG	GCA	CAT	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GCA	GGC	TCC	AGT	GTG	AGC	GTA	TCG	CCC	CCA	CTG
dkik	GCA	CCG	CTG	GCC	CAT	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GC-	--C	TCC	ACT	GTG	AGC	GTT	TCG	CCG	CCG	TTG
dtak	GCG	CCG	CTG	GCA	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GGA	GGC	TCC	AGT	GTG	AGC	GTA	TCG	CCC	CCA	CTG
dpse	GCG	CCG	CTG	GCA	CAT	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GCA	GCC	TCA	ACT	GTG	AGC	GTT	TCG	CCG	CCA	CCG
dwil	GCT	CCC	CTG	GCA	CAT	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAC	GGA	---	---	---	---	---	---	AGC	CTG	ACA	CCG	CCG	CCG
dmoj	GCG	CCG	CTG	GCA	CAT	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCA	GCC	GCC	GTT	TCT	GTG	A--	---	-CG	CCG	CCG	CCG
dalb	GCG	CCC	CTA	GCA	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCC	GCC	GCC	GCC	TCC	GTG	AAT	GTG	TCG	CCG	CCG	CCG
dgri	GCG	CCC	CTG	GCA	CAC	ATG	TGA	GAA	ATG	GCC	AGC	GCC	TAT	GGA	GCC	GCC	GTC	TC-	GCC	GCC	A--	---	CCG	CCG	CCG	CCG

Gene 1

Annotated stop codon

SCR is very abundant in insects





second ORF

T	T	D	F	F	M	P	I	V	D	N	P	F	D.mel
T	T	D	F	F	M	P	I	V	D	D	P	F	D. Bip2
T	V	D	V	F	T	P	I	V	D	D	P	Y	D.Eug2
T	T	D	F	F	Y	P	L	V	E	D	P	Y	D.Tak2
T	V	D	F	F	Y	P	M	V	N	D	P	E	D.Pse3
T	T	D	F	F	Y	P	L	I	D	D	P	Y	D.Rho2

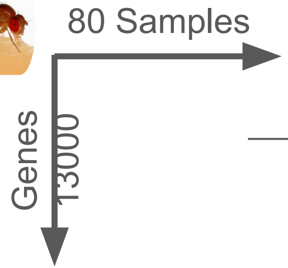
Conservation

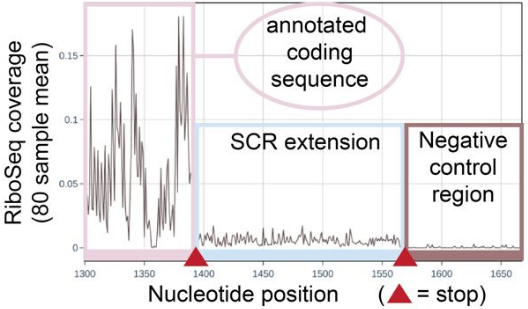
We study **stop codon readthrough (SCR) in insects**

- SCR is particularly abundant in insects
- What is the role of SCR in biological processes?
- In which genes and which tissues is SCR activated?
- **What sequence elements are responsible for it?**

How we detect SCR in ribosome profiling data?

Pipeline 1 - manual selection

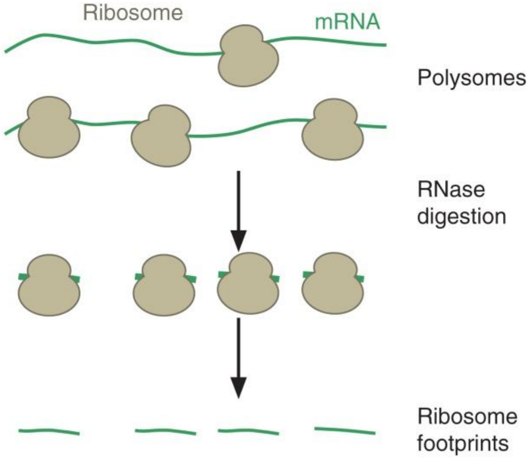




Gish gene

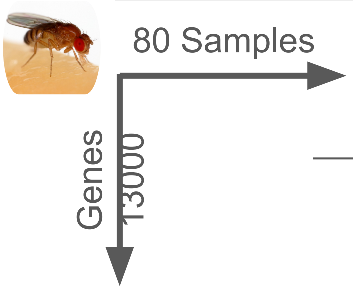
multiple filter criteria

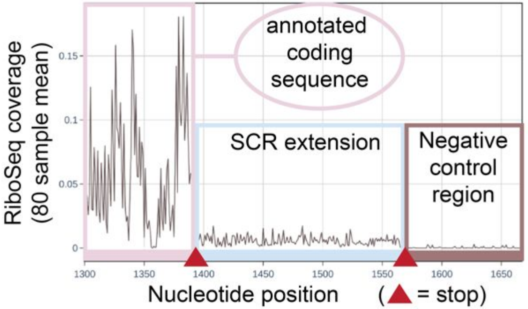
genes with predicted SCR



How we detect SCR in ribosome profiling data?

Pipeline 1 - manual selection





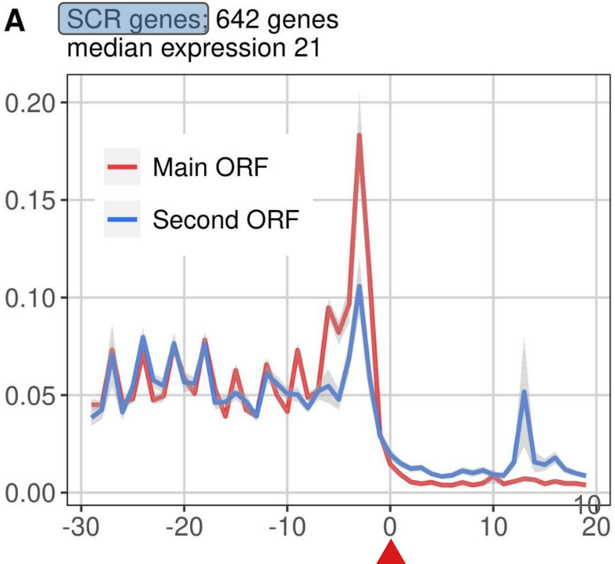
Gish gene

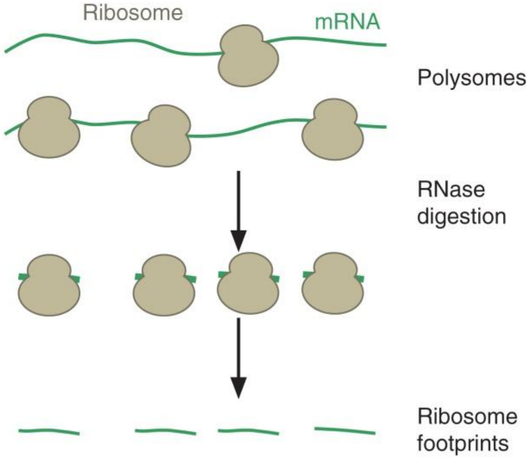
multiple filter criteria

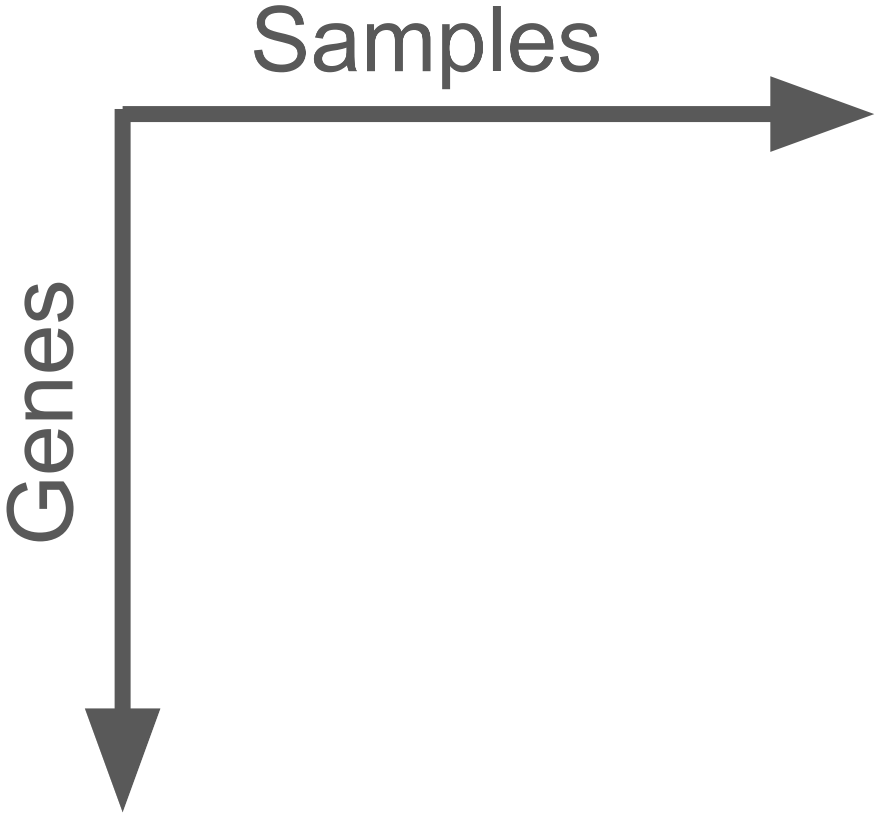
genes with predicted SCR

QC √

3nt periodicity



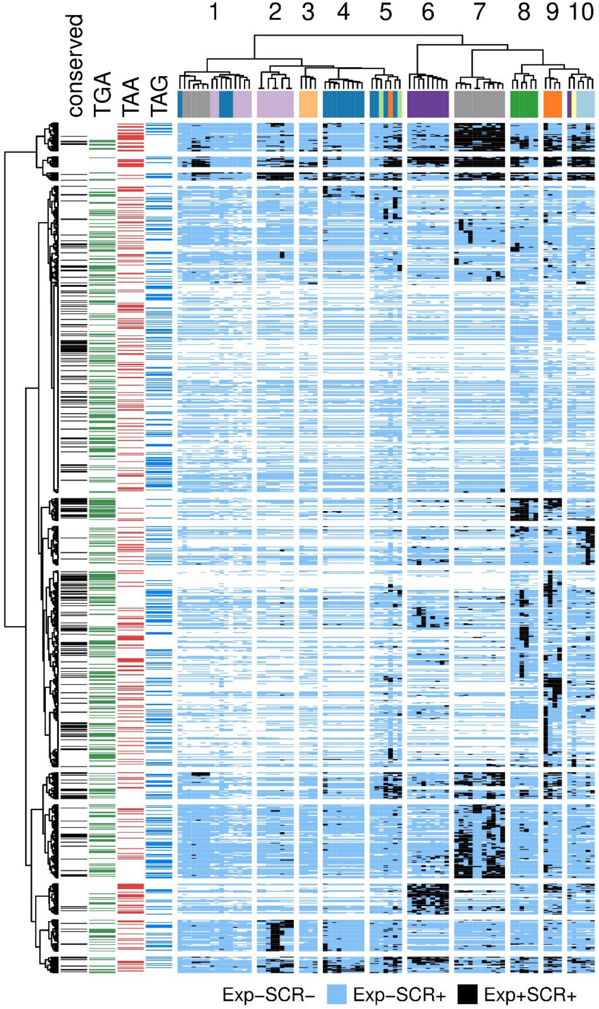




Samples

Genes

The landscape of Drosophila SCR



Sample_category

- 🟦 bodies
- 🟦 early_embryo
- 🟩 heads
- 🟪 larvae
- 🟧 late_embryo
- 🟩 mid_embryo
- 🟪 oocyte
- 🟧 ovaries
- 🟨 pupae
- ⬜ S2_cells

Heatmap legend

- ■ **SCR & expression detected**
- 🟦 **Expression detected**
- □ **Not detected**

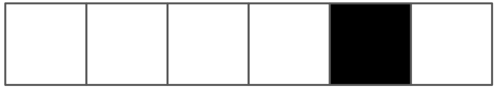
- We detect 1052 genes with SCR (up to 3250 with relaxed criteria)
- Includes SCR genes detected by conservation + others
- SCR is gene specific and tissue/dev.stage specific

The SCR landscape

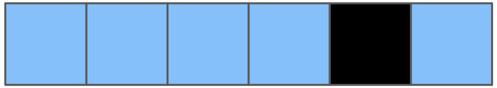
Heatmap legend

■ **SCR & expression detected**
🟦 **Expression detected**
□ **Not detected**

- For others, SCR is present any time the gene is expressed:



- For some genes, SCR is activated in specific tissues/dev.stages:



These depend on SCR factors?

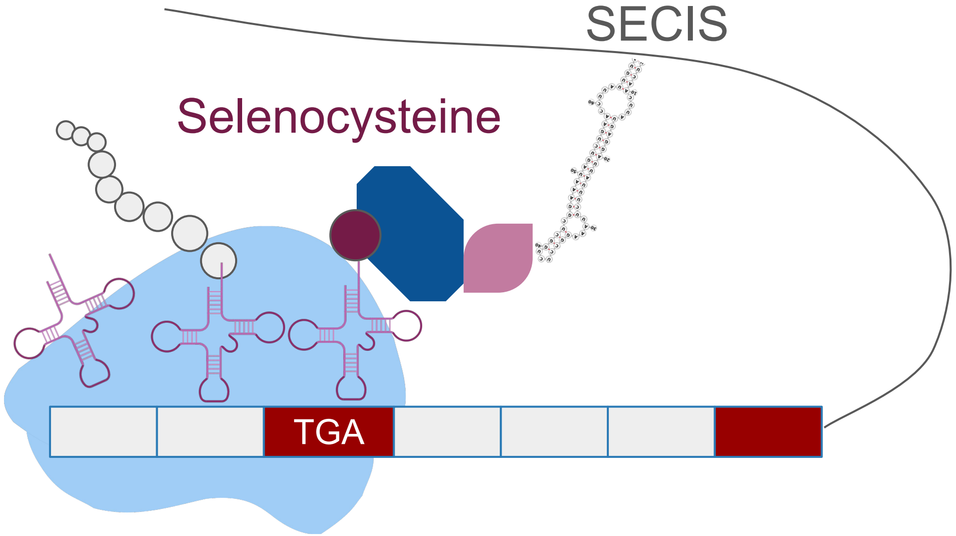
Classification of SCR events

SCR detected by:

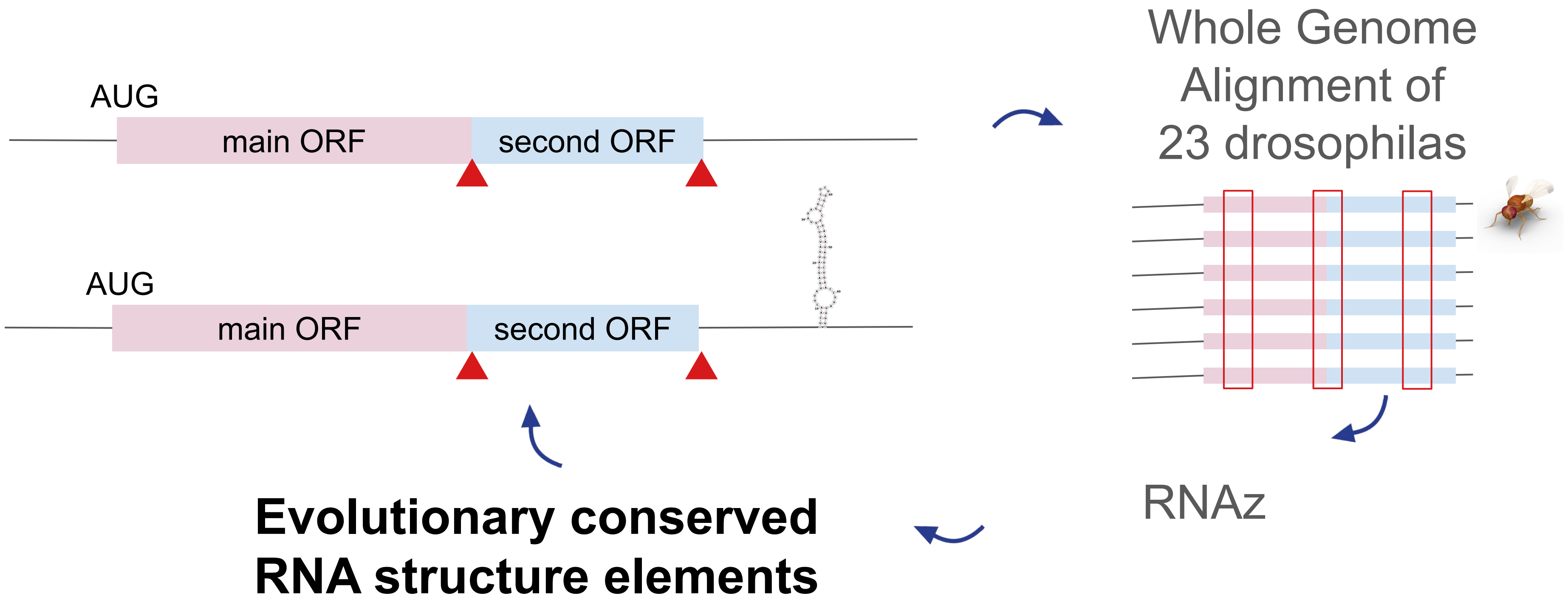
Ribo-Seq +	Conservation +	**Functional SCR**
Ribo-Seq -	Conservation +	**Functional SCR undetected**
Ribo-Seq +	Conservation -	**Accidental or young SCR**
Ribo-Seq -	Conservation -	non SCR

Example of functional structure in 3UTR

A form of recoding= programmed exceptions to the genetic code



Conserved RNA structures in SCR genes

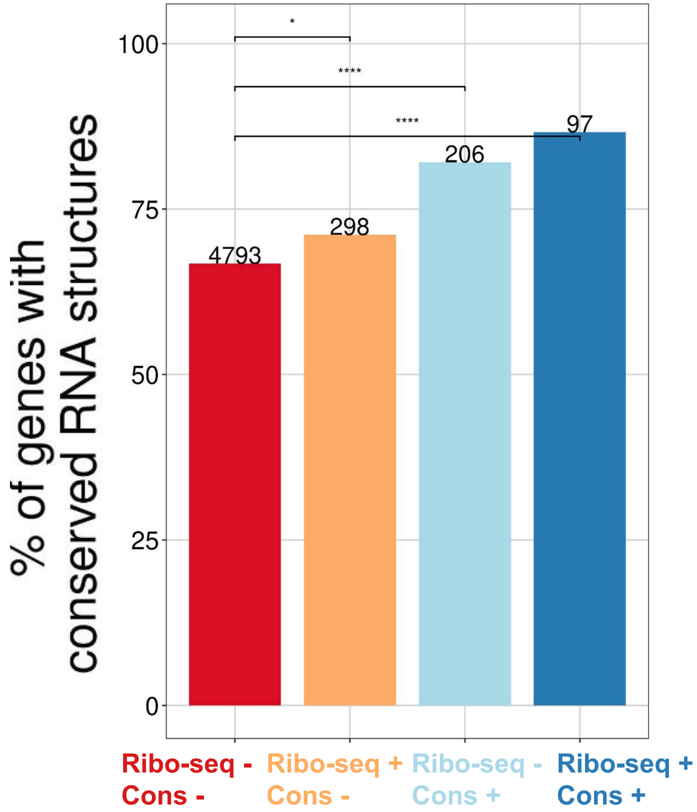


Whole Genome Alignment of 23 drosophilas

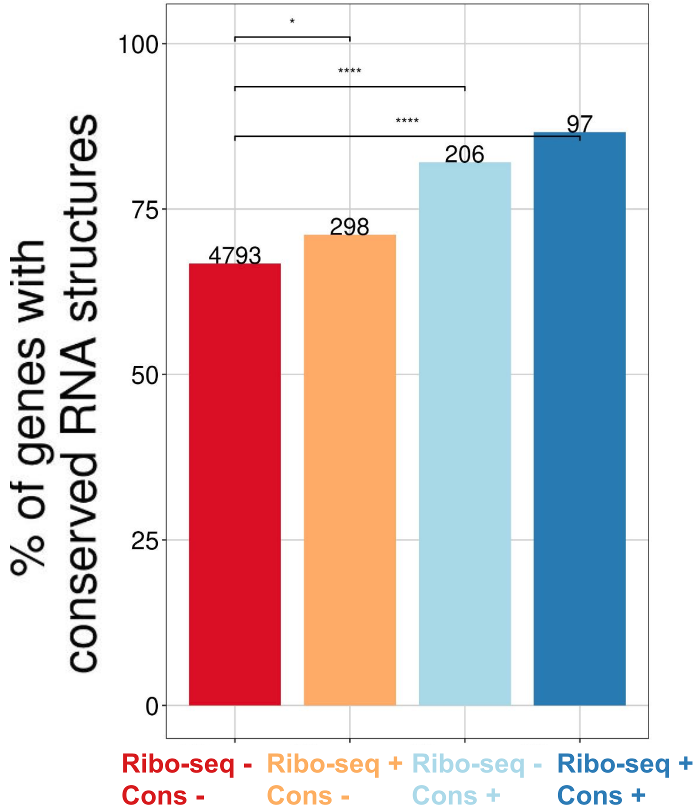
RNAz

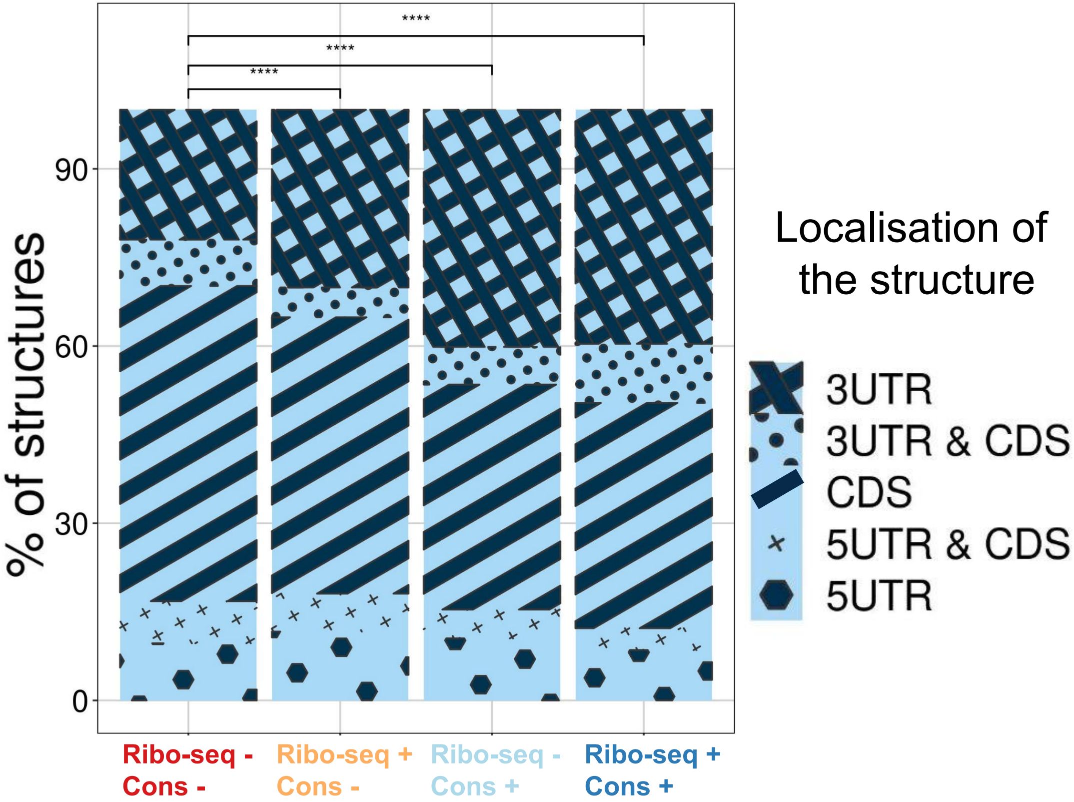
Evolutionary conserved RNA structure elements

SCR genes have more conserved RNA structures, located in the 3' UTR near the stop



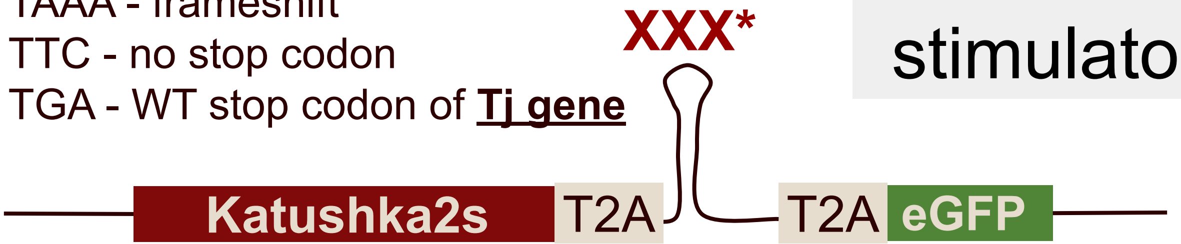
SCR genes have more conserved RNA structures, located in the 3' UTR near the stop

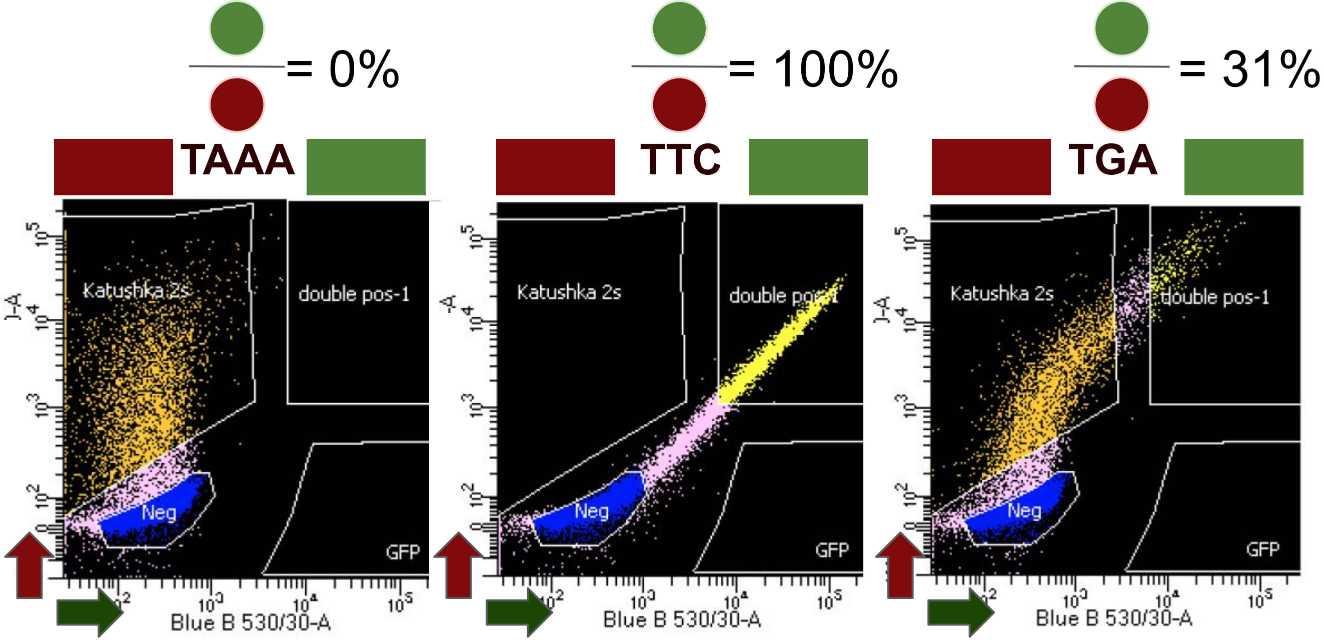




How we assay SCR stimulator sequences

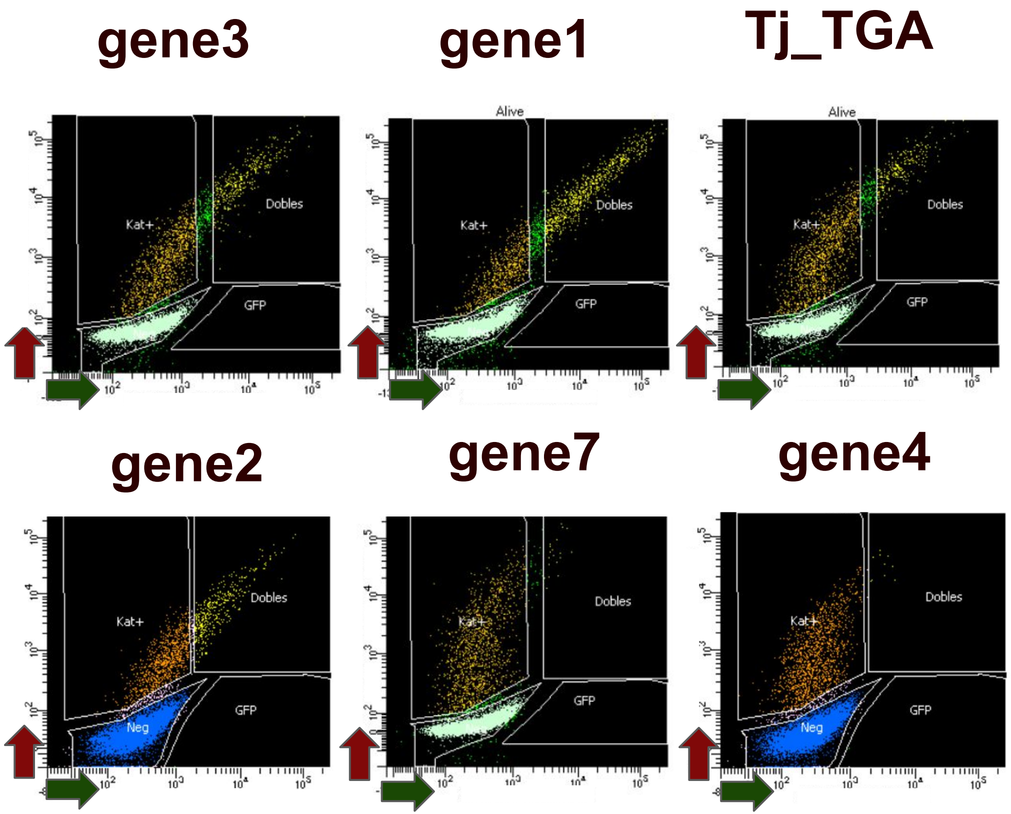
- * TAAA - frameshift
- * TTC - no stop codon
- * TGA - WT stop codon of **Tj gene**



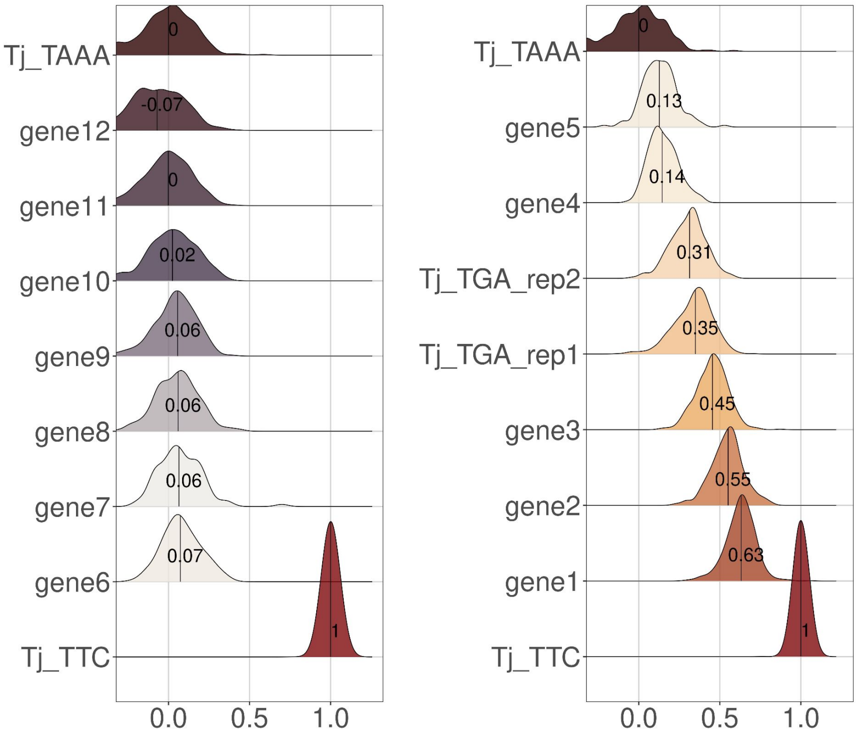


Karki et al. Rodnina, 2022

We discover new SCR candidates!



SCR rate



$$\text{SCR rate} = \frac{\log10(\text{GFP})}{\log10(\text{Katushka2s})}$$

Genes → RNA structures → Mechanism

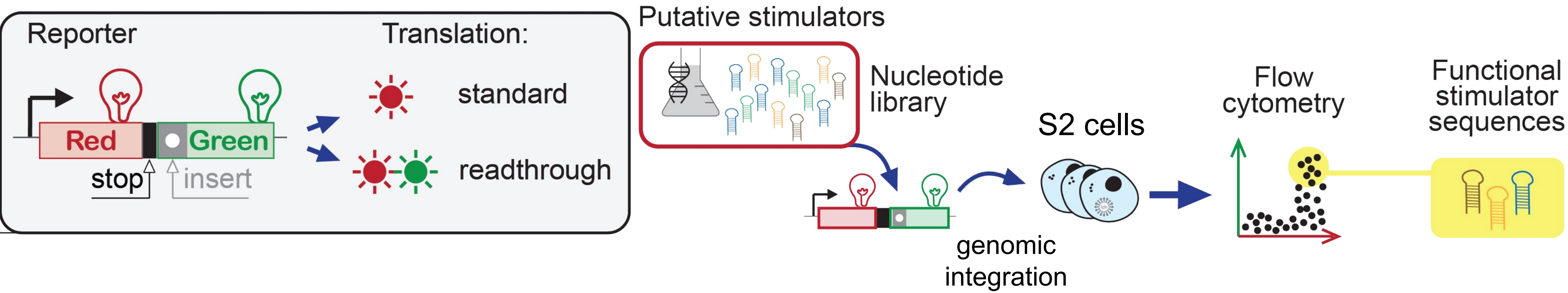
- ~200 genes

- ~400 structures

- tRNA_{sup}
- Rfam
- sequence similarity

Pooled screening of SCR stimulators

Ongoing:



If we can not assign
RNA structure to any known ones,
can we try to group them based on the
structure?

How to compare structures?

Alignment

RNAalifold prediction

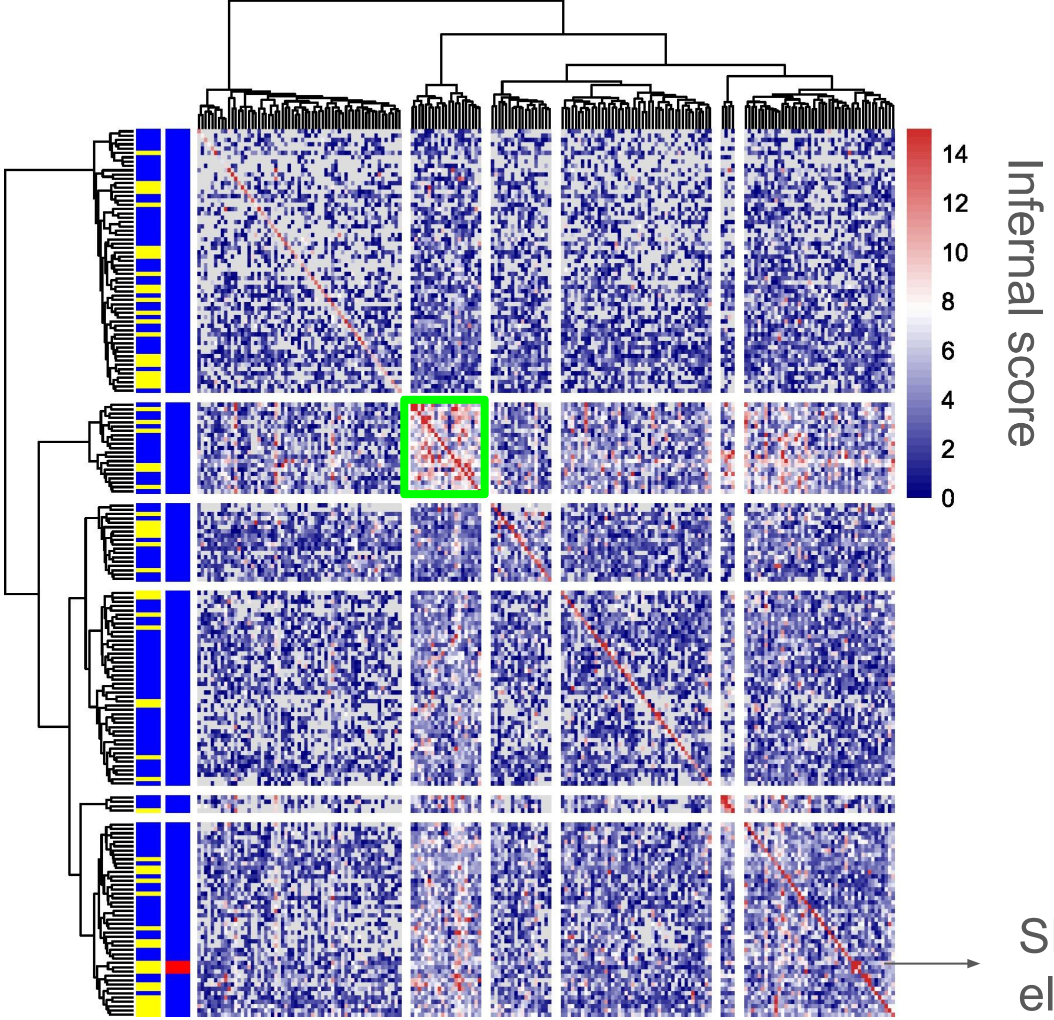
Infernal model building. cmbuild

Infernal model calibration. cmcalibrate

Infernal structure search. cmsearch

database of structures

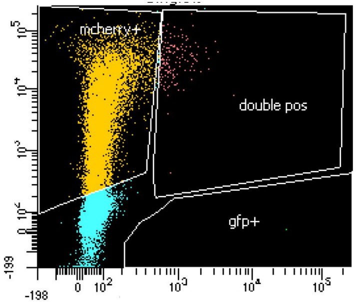
Use score to cluster structures



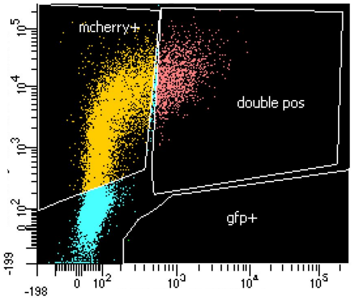
SECIS elements

Our candidates exhibit SCR in mammalian cells

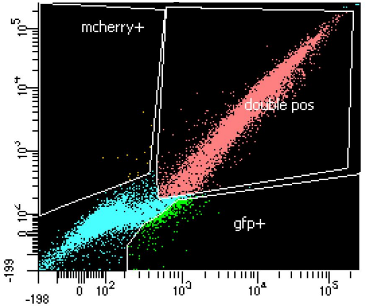
Tj_TAAA



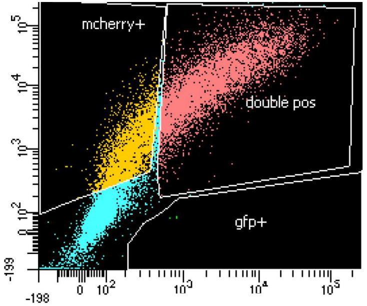
Tj_TGA



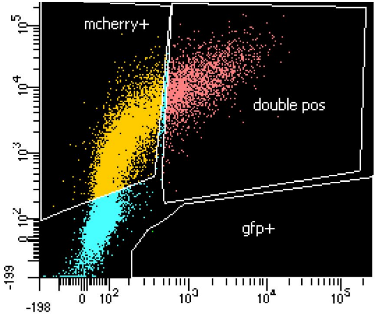
Tj_TTC



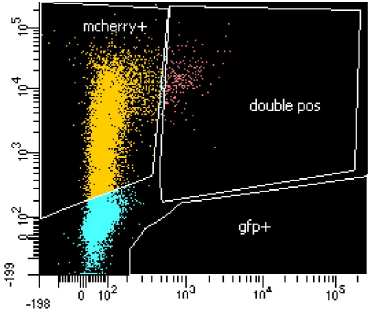
gene1



gene3



gene2



Hek293,
transfection

Conclusions

- Both programmed and accidental SCR are abundant in insects

- The **landscape of SCR** shows gene and tissue/dev.stage specificity

- We found **RNA structures** sufficient for SCR stimulation
    - some SCR stimulators work autonomously
    - others require unknown cell-specific machinery

- Insect SCR is an **essential layer of gene regulation**, and a potential **asset for biotechnology**

MARIOTTI LAB @ UB
COMPARATIVE GENOMICS

www.mariottigenomicslab.com

Marco Mariotti

Max Ticó

Daniel Martín de la Fuente

Ester Muñoz Del Campo

Joan Pallares Albanell

Sergio Sánchez Moragues

Rosina Comas Ghierra

Montserrat Corominas Guiu

Collaborators:

Julie Aspden

Deepika Vasudevan

Pasha Baranov

Open to new one!

AGAUR
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN