

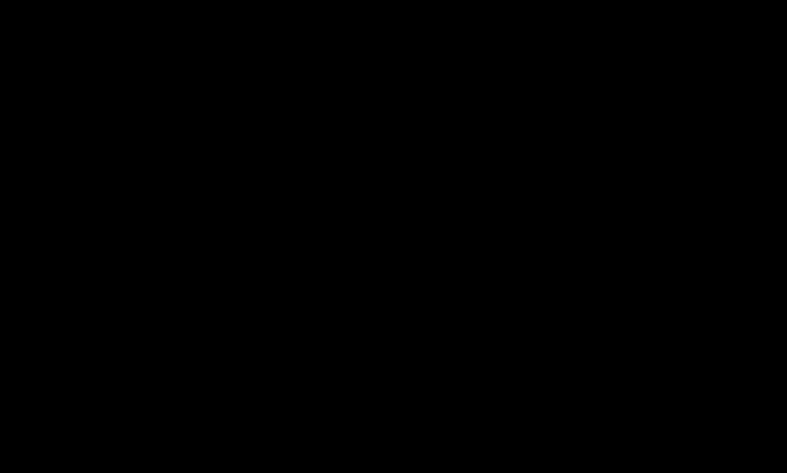


Genetic mosaicism in the *Drosophila* intestine –
from somatic mutations to endogenous retrotransposon activity

Kasia Siudeja
Team Nuclear Dynamics and Repetitive DNA
I2BC, Paris-Saclay



Psiloritis 2006

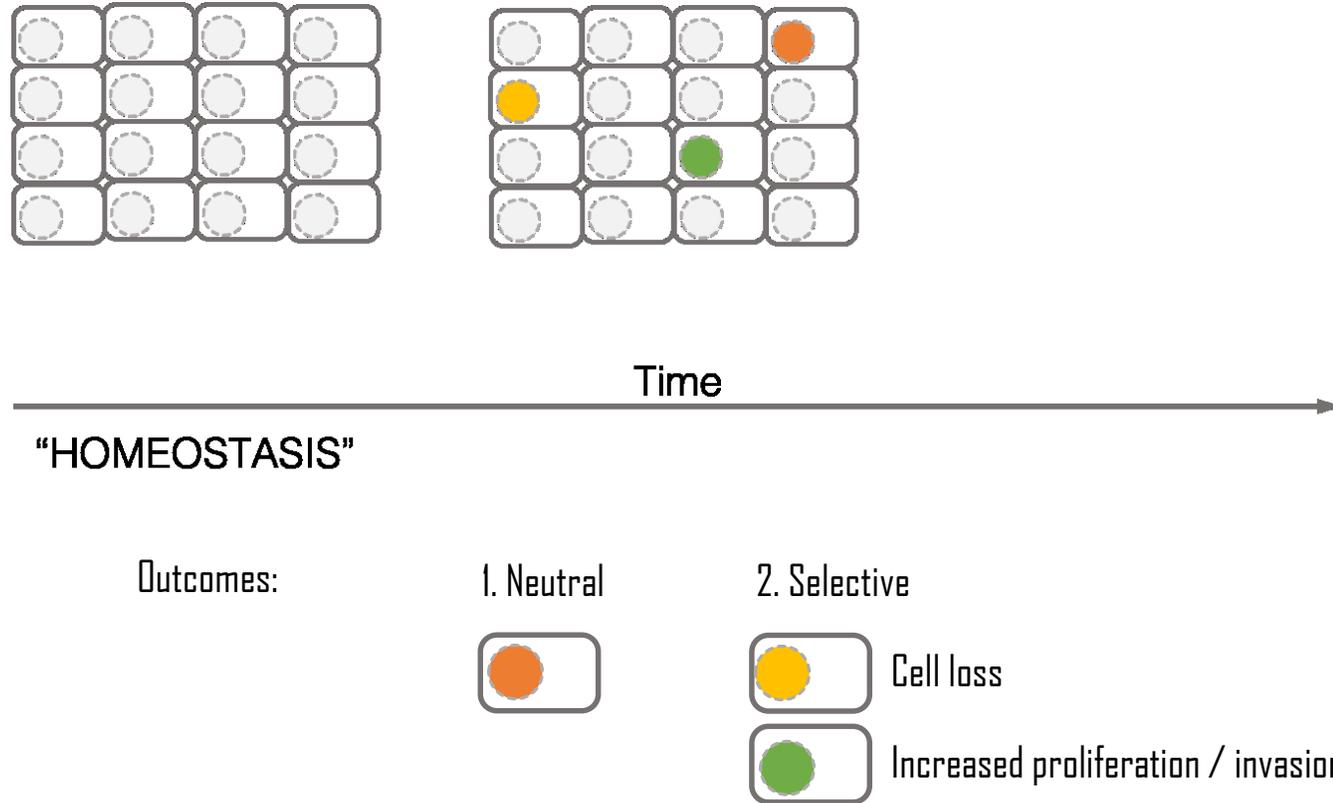


Psiloritis 2006



Ida 2021

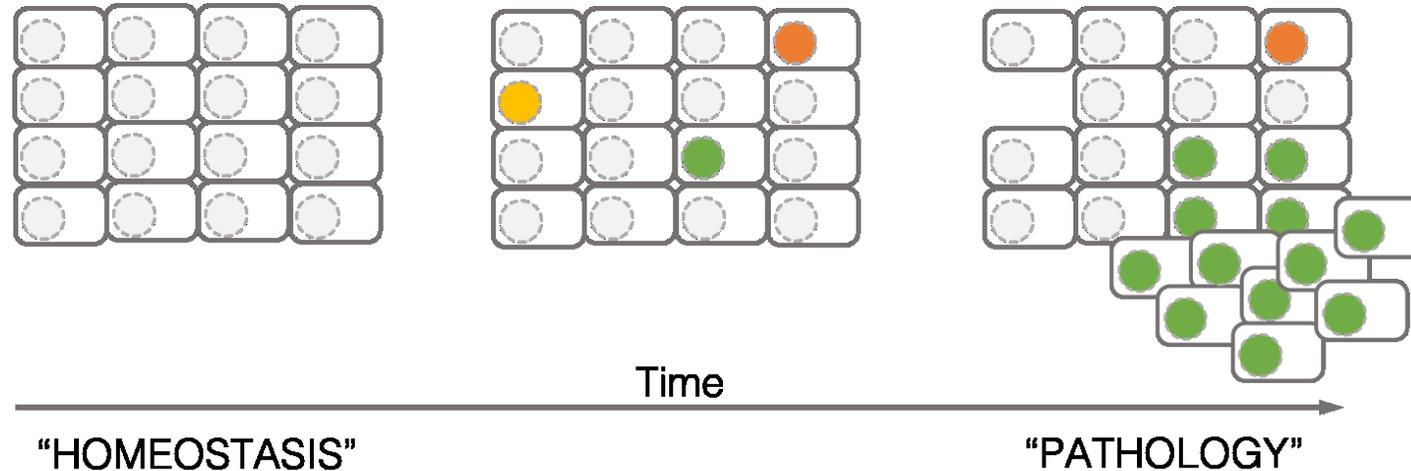
Genome plasticity influences cell function *in vivo*



Selected refs:

- Blood (Laurie, Nat Gen, 2012; Jacobs, Nat Gen, 2012; Genovese, NEJM, 2014; Jaiswal, NEJM, 2014; Lee-Six, Nature, 2018)
- Skin (Martincorena, Science, 2015)
- Esophagus (Martincorena, Science 2018)
- Liver (Bruner, Nature, 2019)
- Colon (Lee-Six, Nature, 2019)

Genome plasticity influences cell function *in vivo*



Outcomes:

1. Neutral



2. Selective



Cell loss

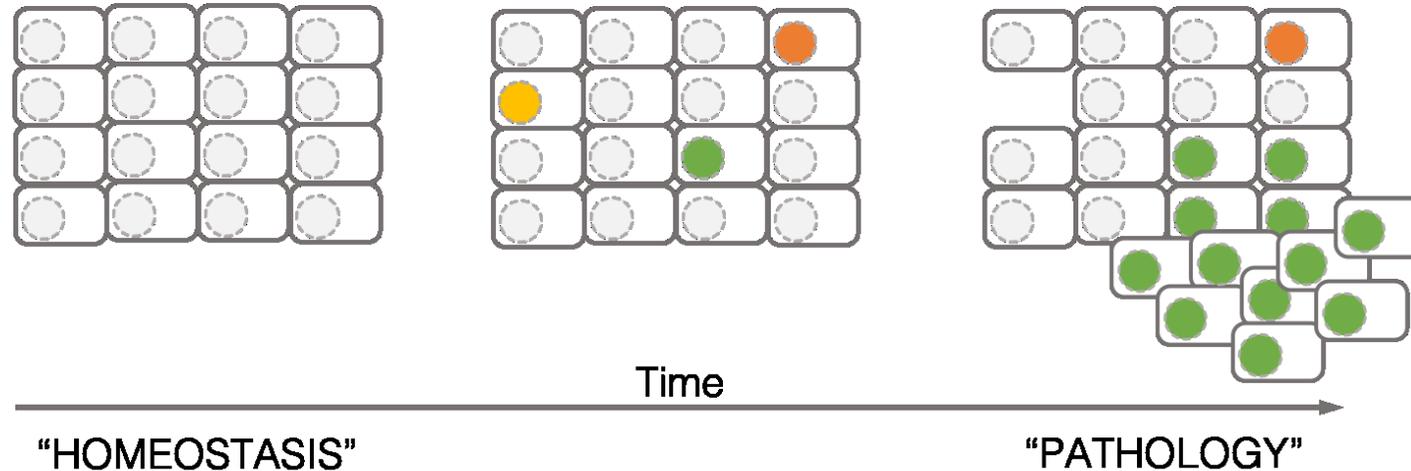


Increased proliferation / invasion

Selected refs:

- Blood (Laurie, Nat Gen, 2012; Jacobs, Nat Gen, 2012; Genovese, NEJM, 2014; Jaiswal, NEJM, 2014; Lee-Six, Nature, 2018)
- Skin (Martincorena, Science, 2015)
- Esophagus (Martincorena, Science 2018)
- Liver (Bruner, Nature, 2019)
- Colon (Lee-Six, Nature, 2019)

Genome plasticity influences cell function *in vivo*



Outcomes:

1. Neutral



2. Selective



Cell loss



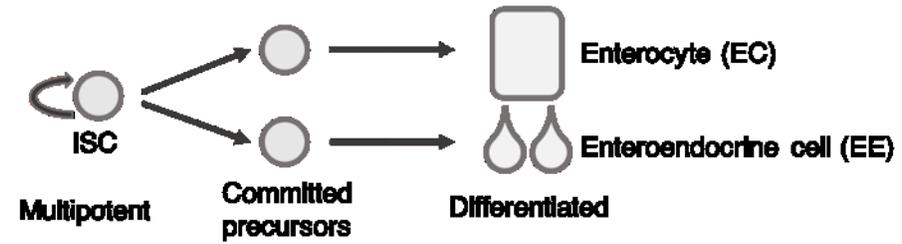
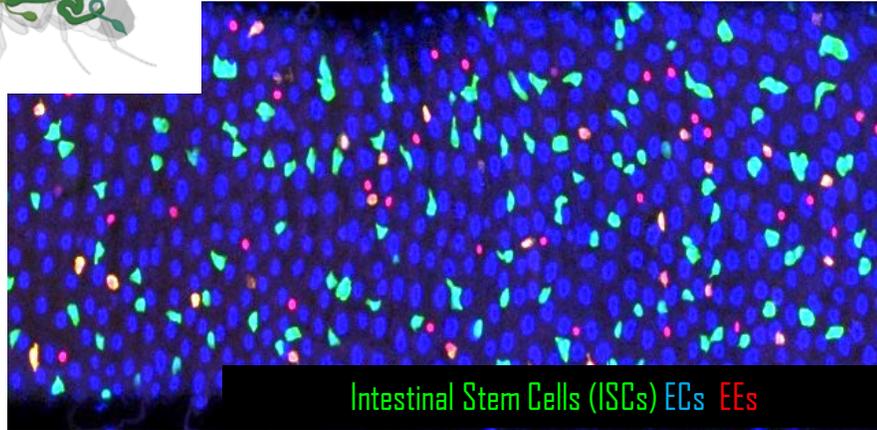
Increased proliferation / invasion

Mechanisms ?

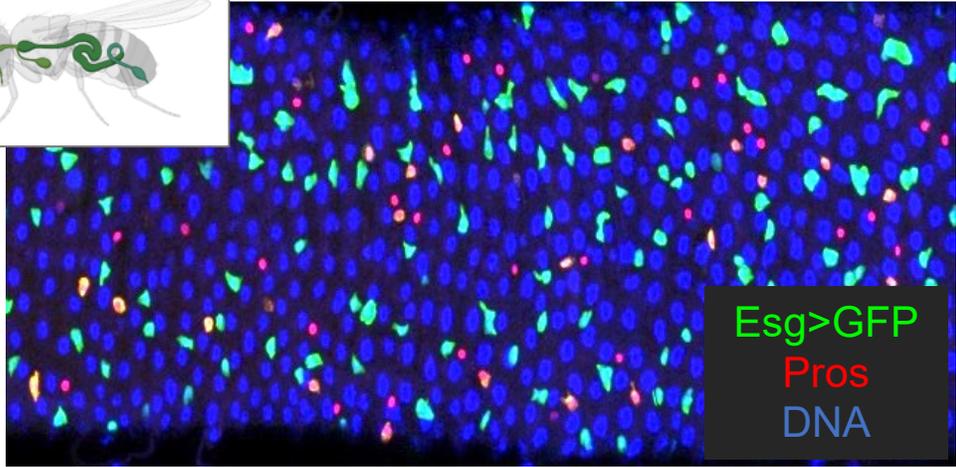
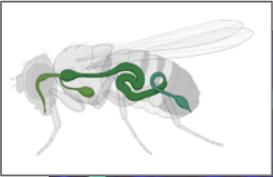
Selected refs:

- Blood (Laurie, Nat Gen, 2012; Jacobs, Nat Gen, 2012; Genovese, NEJM, 2014; Jaiswal, NEJM, 2014; Lee-Six, Nature, 2018)
- Skin (Martincorena, Science, 2015)
- Esophagus (Martincorena, Science 2018)
- Liver (Bruner, Nature, 2019)
- Colon (Lee-Six, Nature, 2019)

The *Drosophila* intestine – model system



Mapping somatic genetic variation *in vivo*

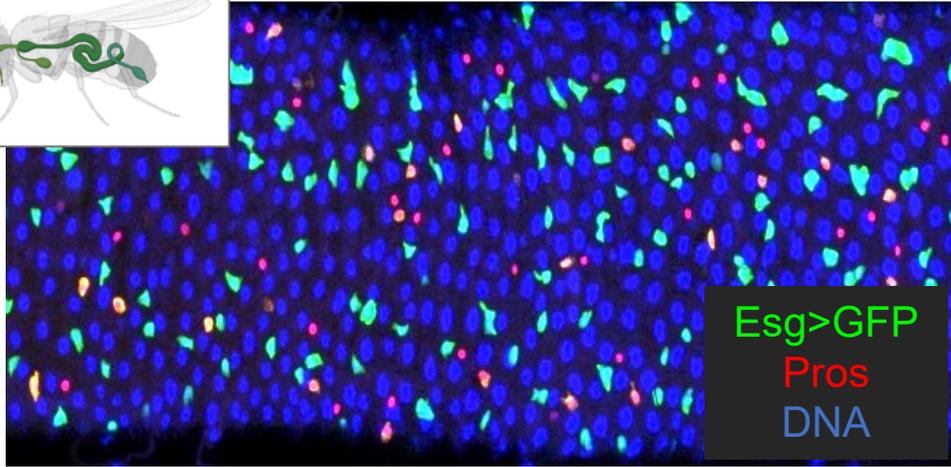
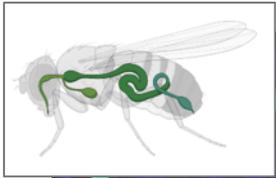


Aging

Somatic mutation

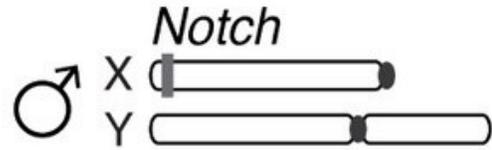
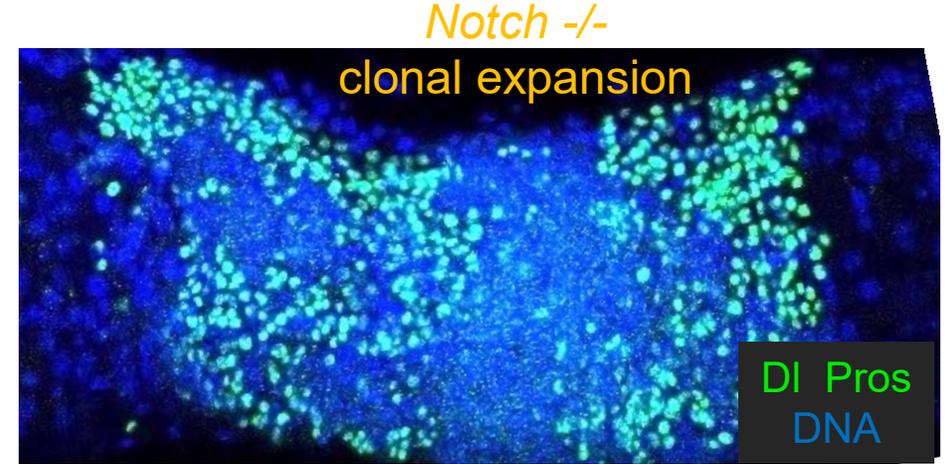


Mapping somatic genetic variation *in vivo*

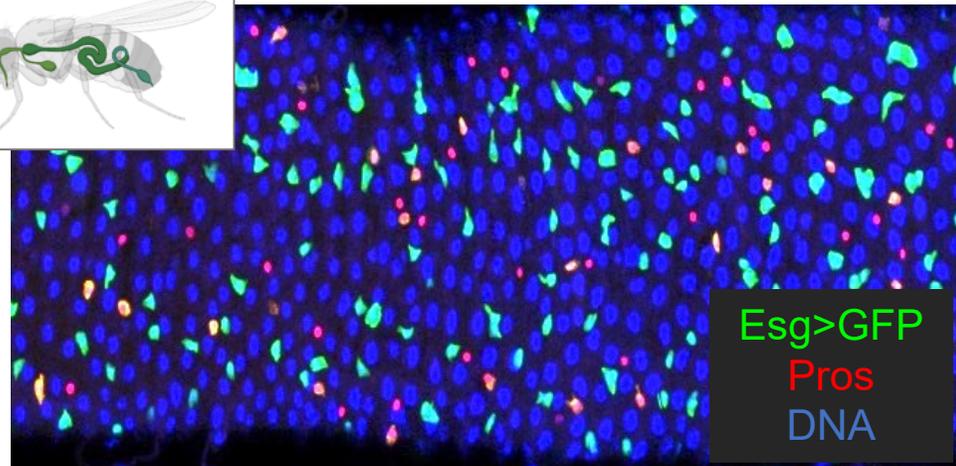
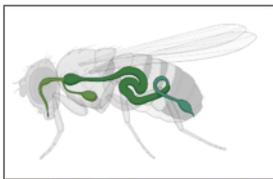


Aging
Somatic mutation

Notch tumor
suppressor
inactivation

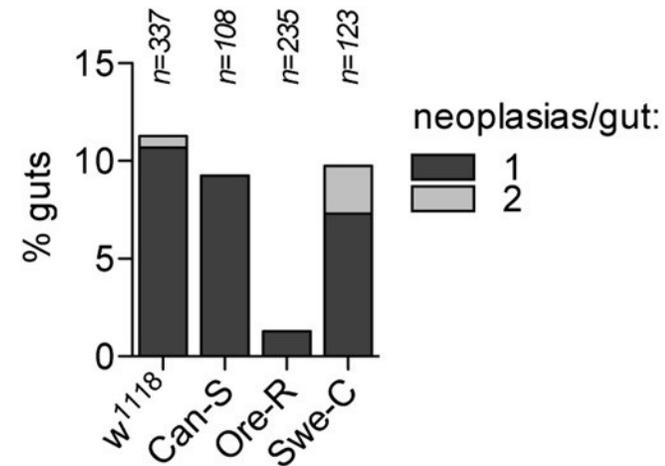
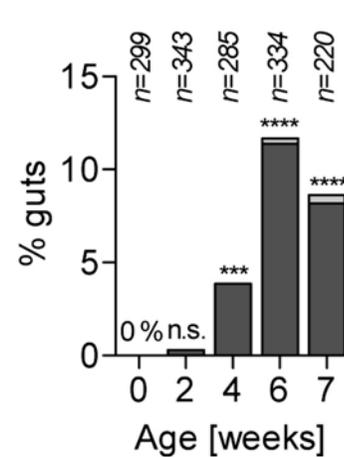
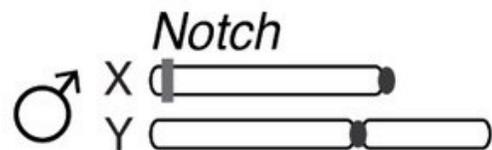
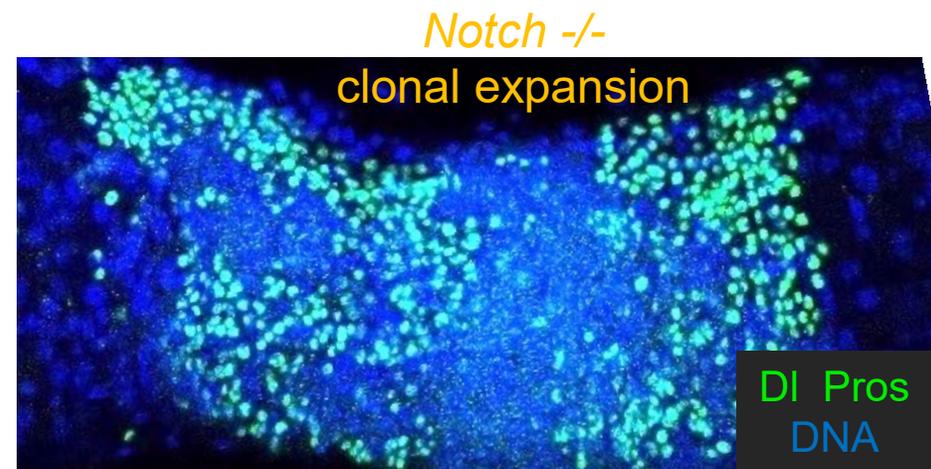


Mapping somatic genetic variation *in vivo*

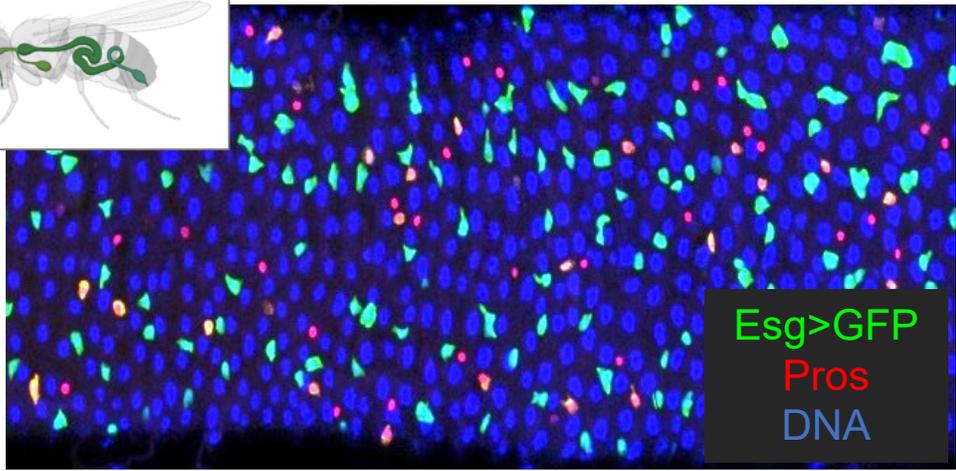
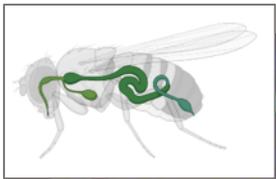


Aging
Somatic mutation

Notch tumor suppressor inactivation



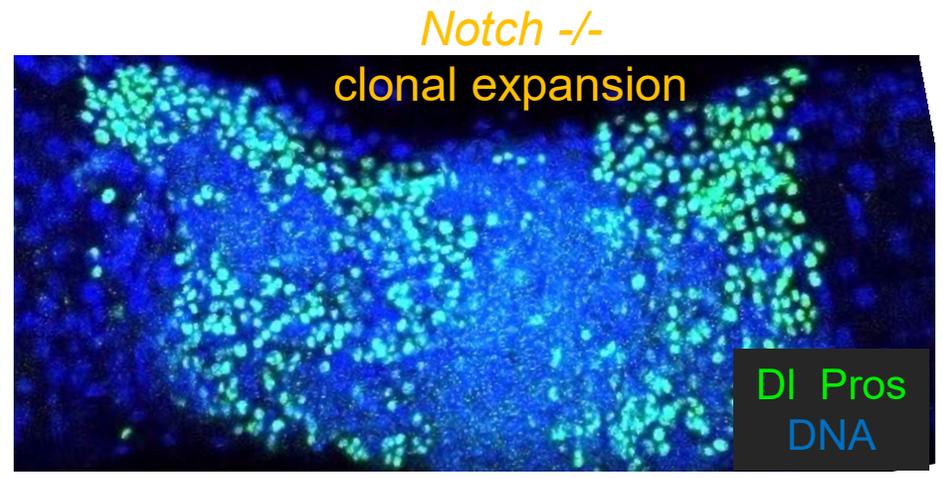
Mapping somatic genetic variation *in vivo*



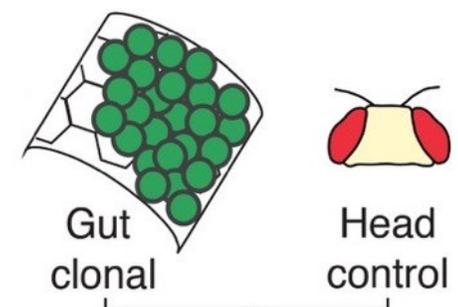
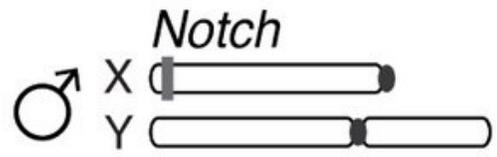
Esg>GFP
Pros
DNA

Aging
Somatic mutation

Notch tumor suppressor inactivation

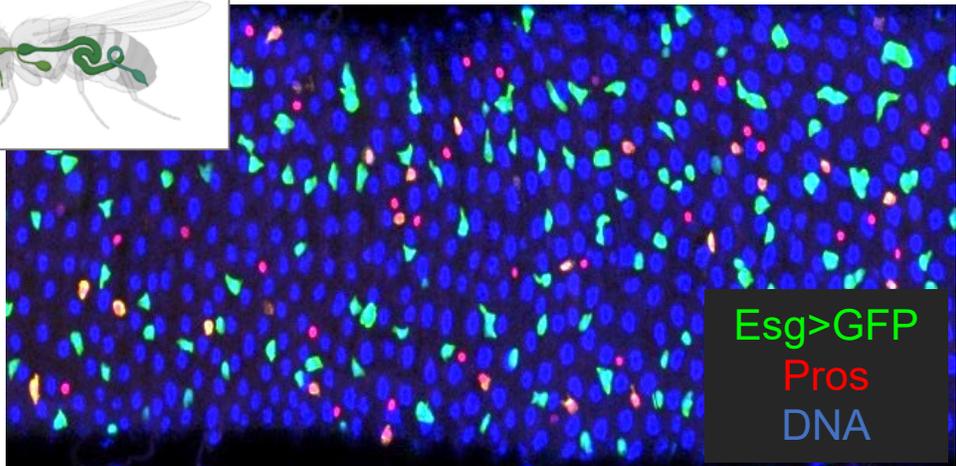
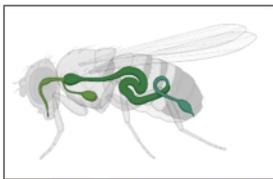


DI Pros
DNA



Whole genome sequencing
(Illumina, 2x100bp or 2x150bp)

Mapping somatic genetic variation *in vivo*

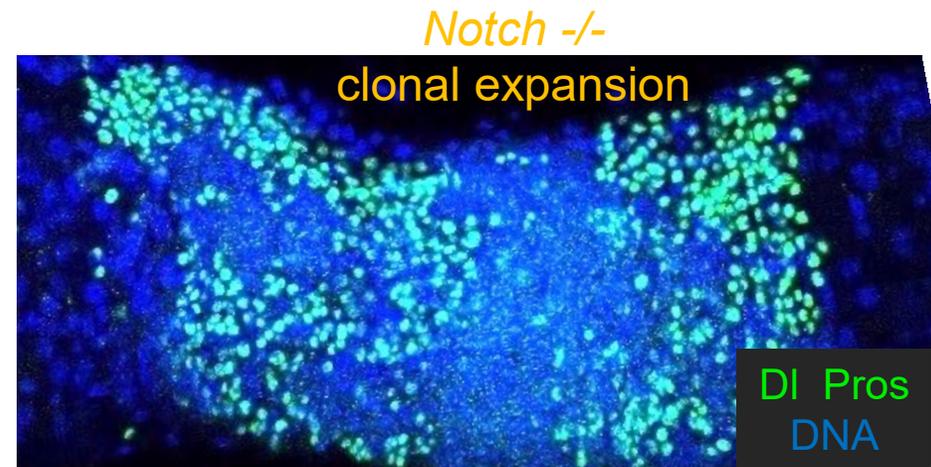


Esg>GFP
Pros
DNA

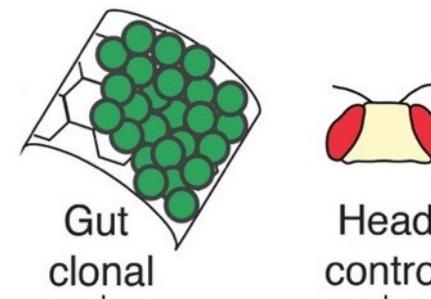
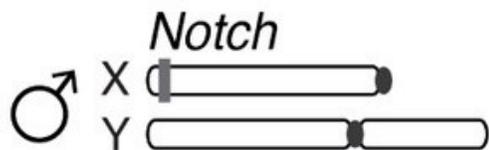
Aging

Somatic mutation

Notch tumor
suppressor
inactivation



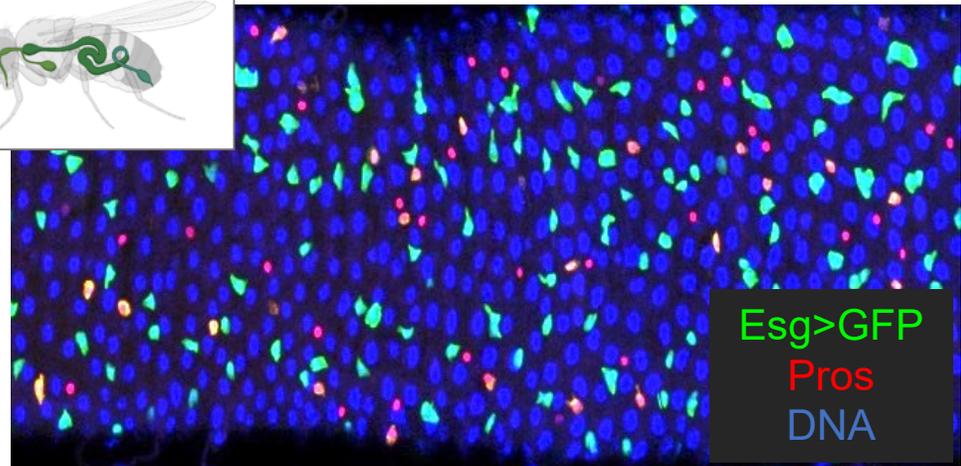
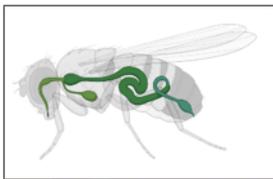
DI Pros
DNA



- Structural variants (deletions / complex rearrangements)
- Point mutations
- Loss of heterozygosity (LOH)
- ...

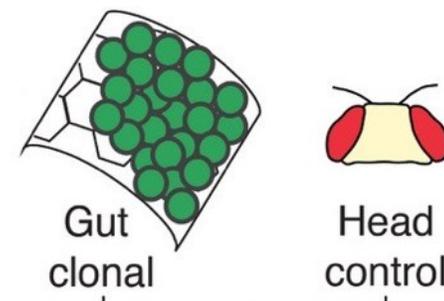
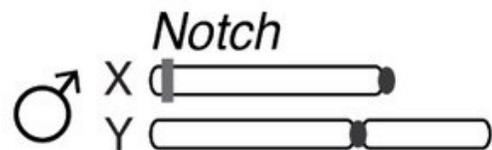
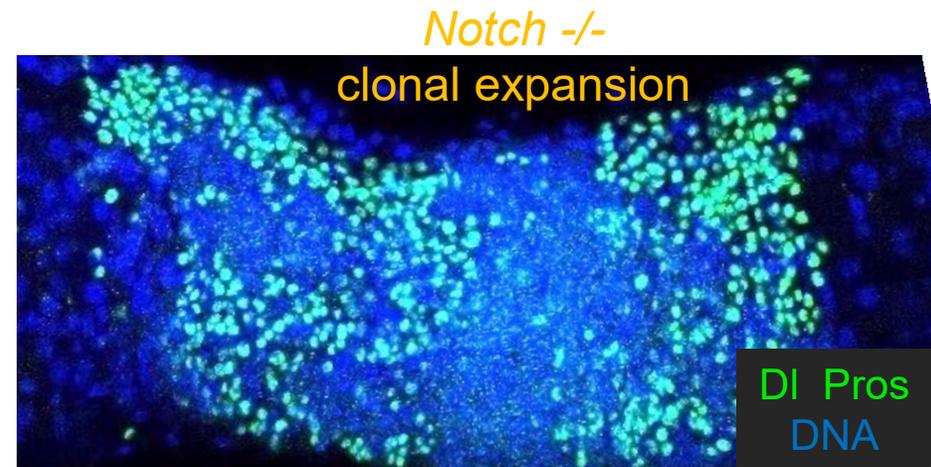
Whole genome sequencing
(Illumina, 2x100bp or 2x150bp)

Mapping somatic genetic variation *in vivo*



Aging
Somatic mutation

Notch tumor suppressor inactivation



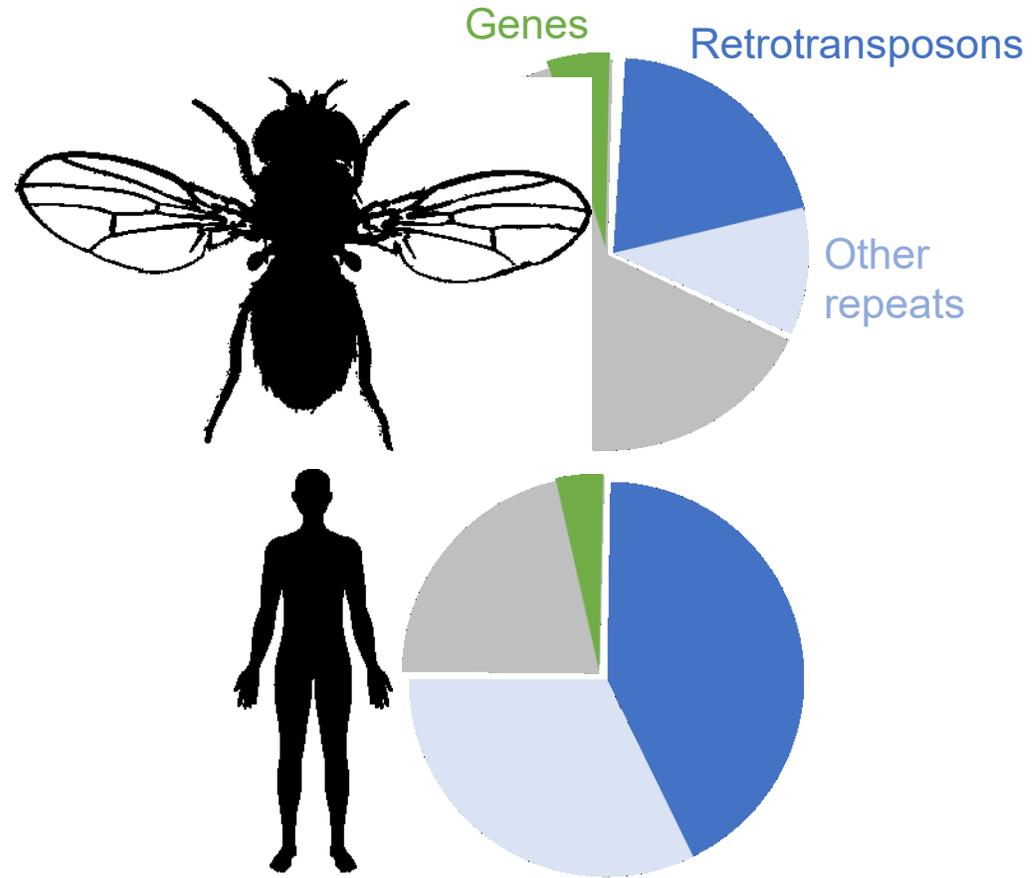
Whole genome sequencing (Illumina, 2x100bp or 2x150bp)

de novo insertions of transposable elements

- Structural variants (deletions / complex rearrangements)
- Point mutations
- Loss of heterozygosity (LOH)
- ...

Retrotransposons - mobile genetic elements

Genome:



Transposable elements

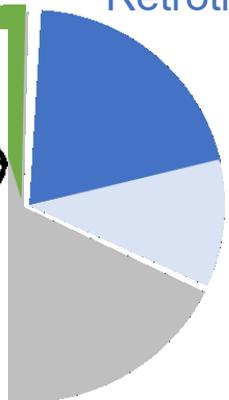
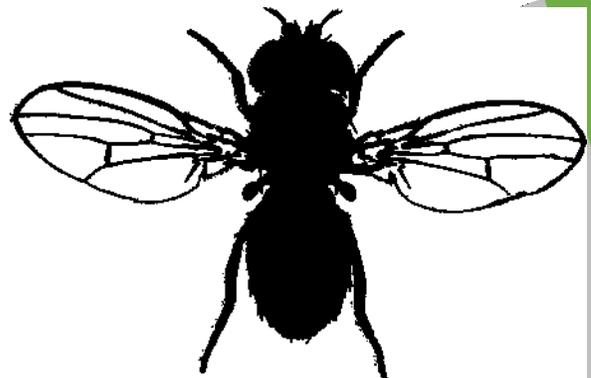
Class 1
Retrotransposons

Class 2
DNA transposons

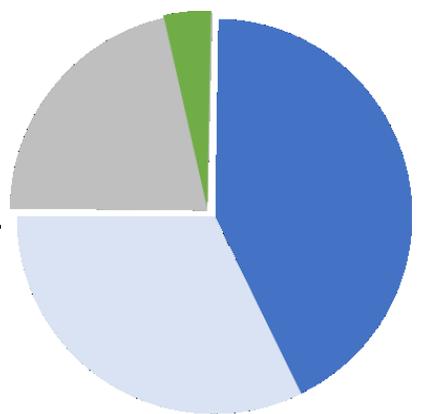
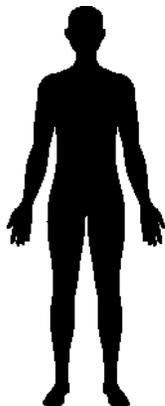
Retrotransposons - mobile genetic elements

Genome:

Genes
Retrotransposons



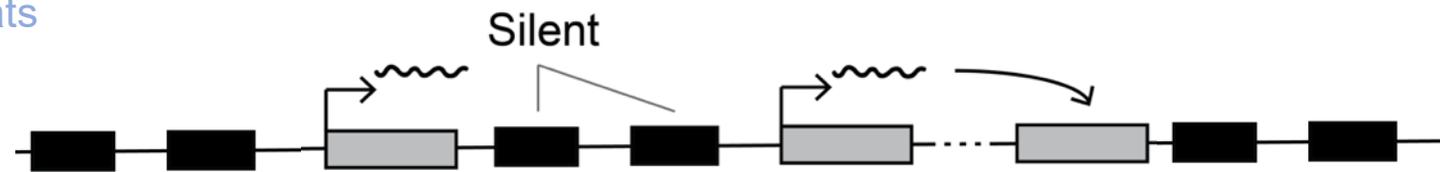
Other repeats



Transposable elements

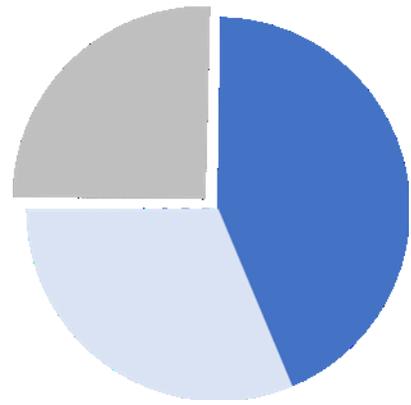
Class 1
Retrotransposons

Class 2
DNA transposons

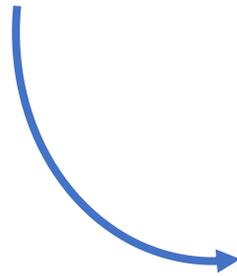


- Able to propagate in the host genomes
- Present in many copies
- Mostly repressed pre- and post-transcriptionally
- BUT:
- Some transcribed and a fraction able to self-propagate

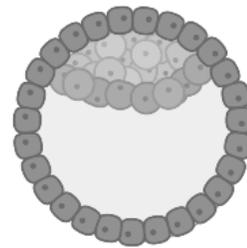
Retrotransposons in the soma



Retrotransposons



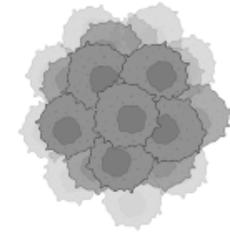
Early development



Neuronal lineages



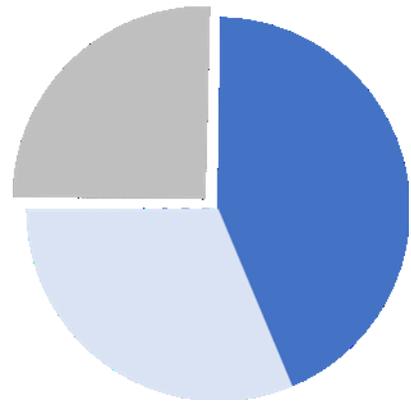
Cancers



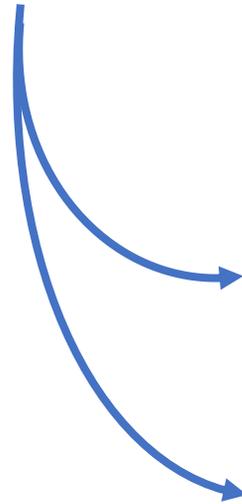
Aging



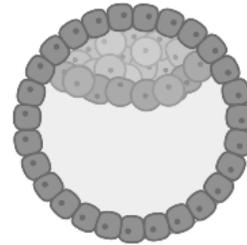
Retrotransposons in the soma



Retrotransposons



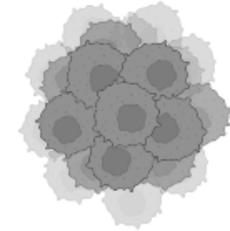
Early development



Neuronal lineages



Cancers

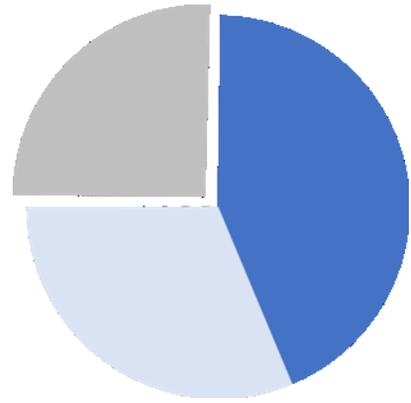


Aging



Activity reported in diverse tissues

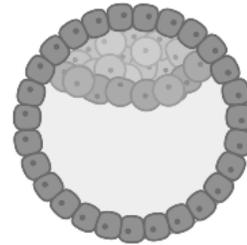
Retrotransposons in the soma



Retrotransposons

?

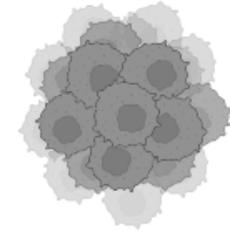
Early development



Neuronal lineages



Cancers



Aging

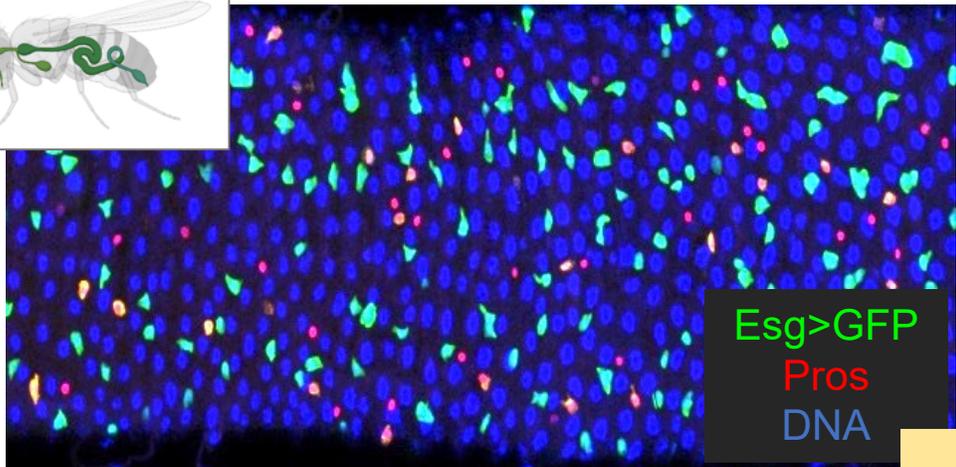
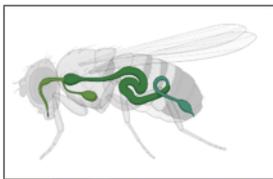


Activity reported in diverse tissues

Open questions:

Mechanisms of activity?
Impact on somatic lineages?

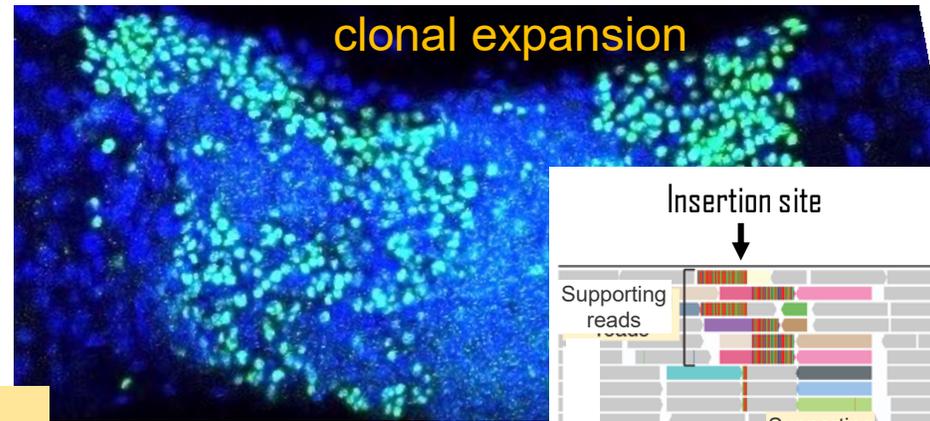
Retrotransposon insertions in *Notch* in spontaneous neoplasia



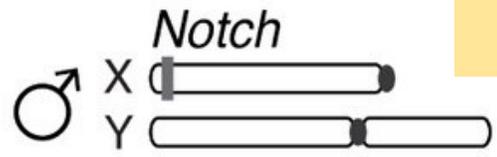
Esg>GFP
Pros
DNA

Aging
Somatic mutation

Notch tumor suppressor inactivation



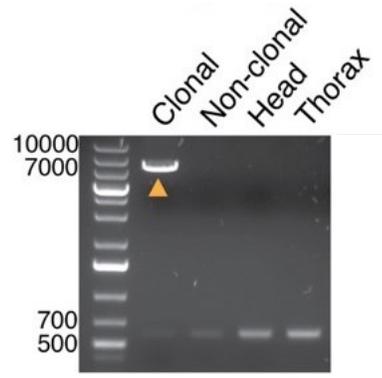
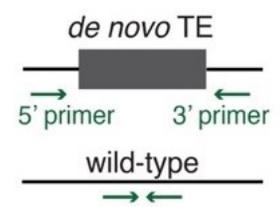
Notch -/-
clonal expansion



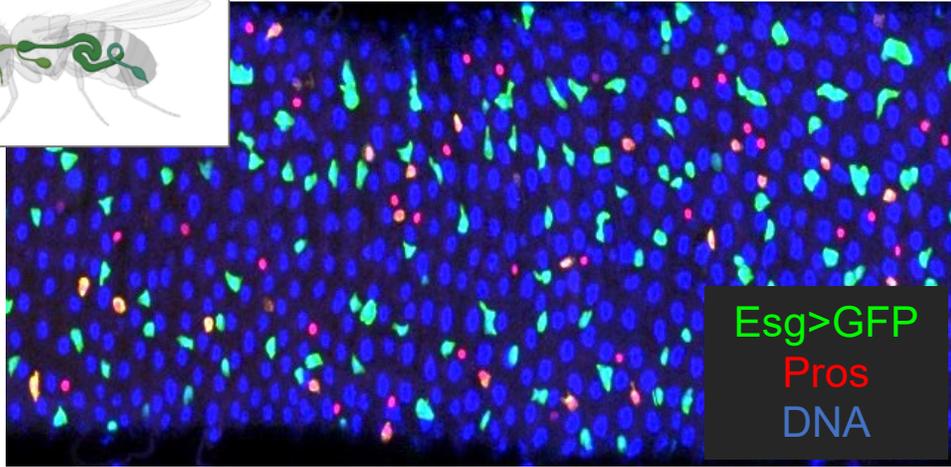
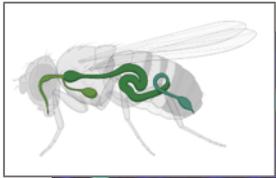
de novo retrotransposon insertions



Full-length PCR:

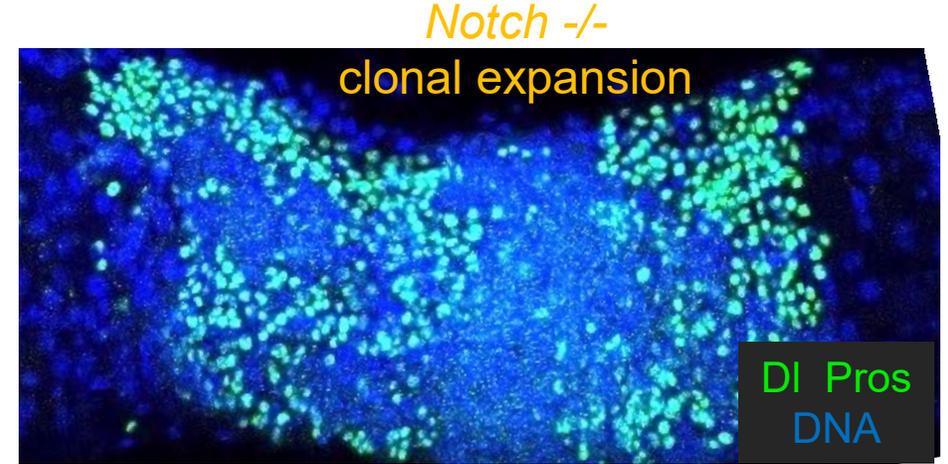


Short- and long-read DNA-seq – revealing genome-wide TE mobility



Aging
Somatic mutation

Notch tumor
suppressor
inactivation



Long-read (ONT) DNA-seq

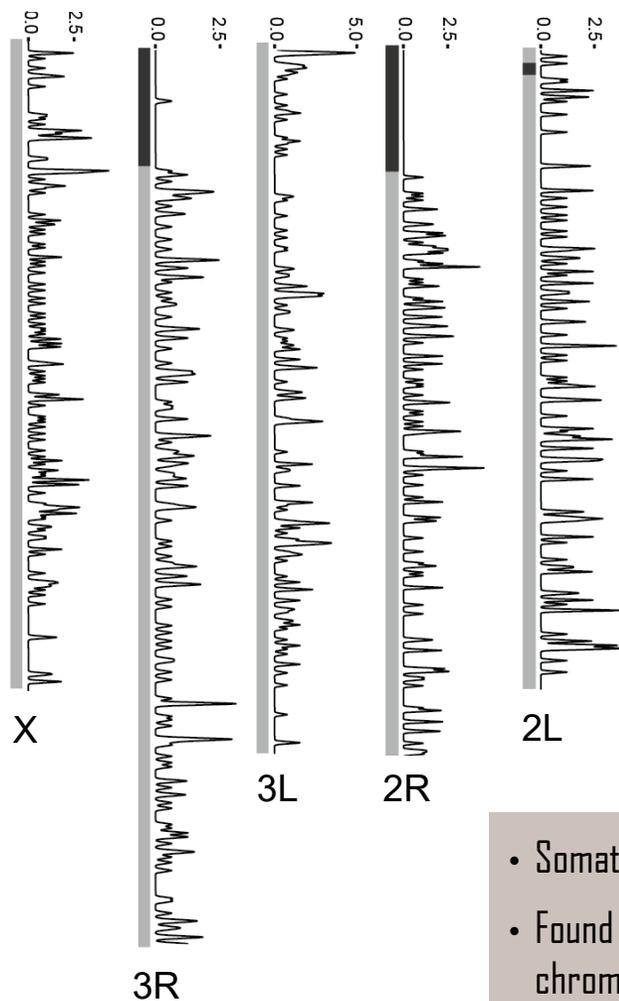


Short-read (Illumina) DNA-seq

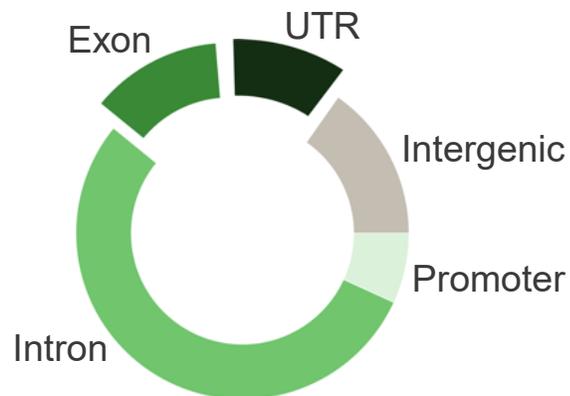


Short- and long-read DNA-seq – revealing genome-wide TE mobility

• Insertion density



• Insertion distribution



• Gene hits

Hits	Gene	Function
Clonal samples (short-read seq):		
5	<i>pointed</i>	EGFR pathway
3	<i>puckered</i>	JNK pathway
3*	<i>Notch</i>	Notch pathway
2	<i>scribble</i>	Epithelial polarity
1	<i>kuzbanian</i>	Notch pathway
1	<i>neuralized</i>	Notch pathway
1	<i>Stat92E</i>	JAK/STAT pathway
1	<i>frizzled</i>	Wnt pathway
1	<i>Nipped-A</i>	ISC proliferation
1	<i>kismet</i>	Chromatin organization
1	<i>osa</i>	Chromatin organization
1	<i>Atg8a</i>	Autophagy
1	<i>robo1</i>	Lineage specification
Pooled guts (long-read seq):		
1	<i>EGFR</i>	EGFR pathway
1	<i>InR</i>	Insulin pathway
1	<i>Myc</i>	ISC proliferation
1	<i>Pvf1</i>	VEGF pathway
1	<i>Pvf2</i>	VEGF pathway
1	<i>Pvf3</i>	VEGF pathway
1	<i>Piezo</i>	Lineage specification
1	<i>pros</i>	Lineage specification
* Insertions with a valid TSD. 1 additional <i>Notch</i> insertion without TSD is reported in Fig1		

- Somatic TE insertion sites spread genome-wide
- Found preferentially in genic regions and open, transcriptionally active chromatin

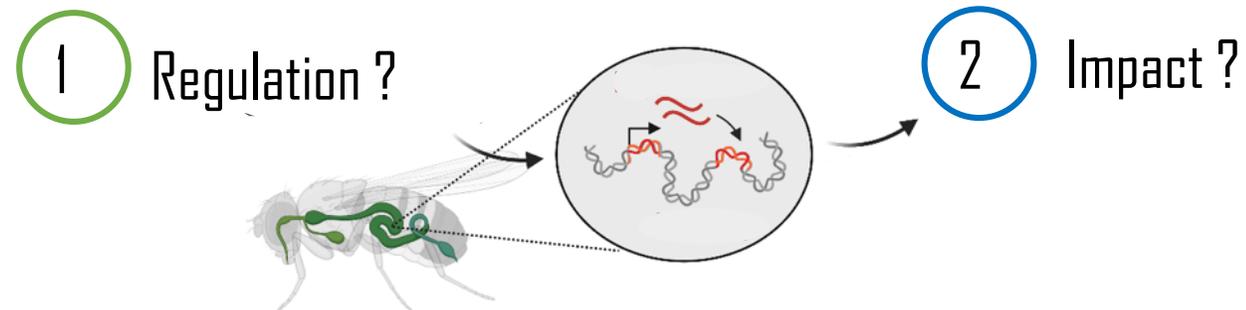
Pre-insertion
site choice

+

Post-insertion
tissue selection

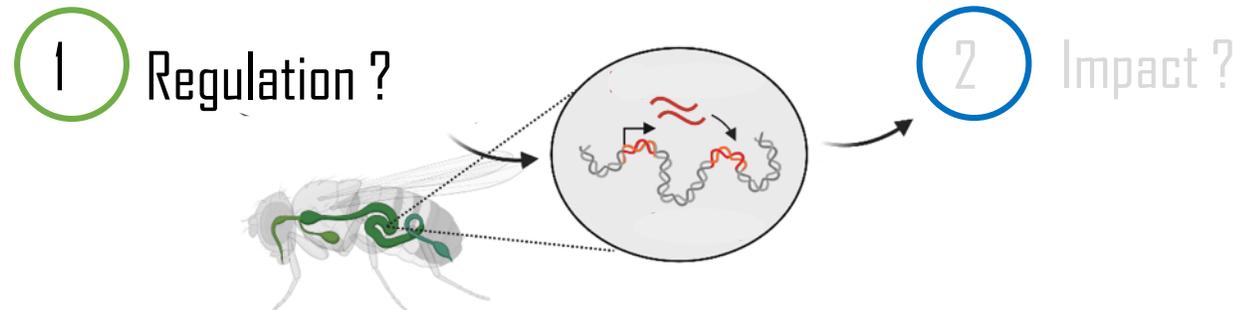
Conclusion 1

- Revealed prevalent genome instability and retrotransposition occurring in the fly midgut
- TE insertions enriched in open chromatin, genic or regulatory regions
- May lead to phenotypic consequences
(*Notch* inactivation -> neoplasia formations)
- Ongoing/future questions:

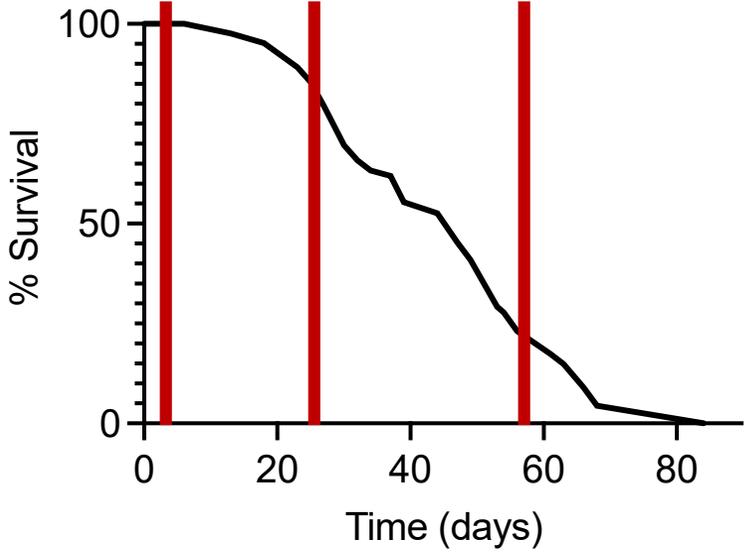
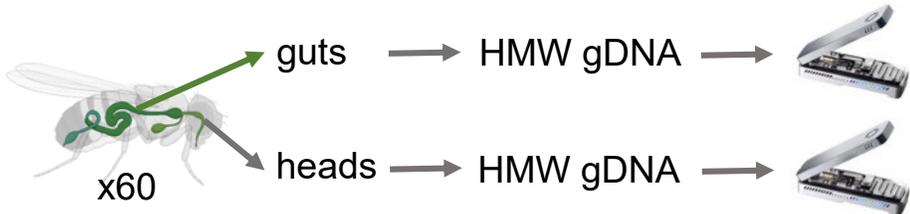


Conclusion 1

- Revealed prevalent genome instability and retrotransposition occurring in the fly midgut
- TE insertions enriched in open chromatin, genic or regulatory regions
- May lead to phenotypic consequences
(*Notch* inactivation -> neoplasia formations)
- Ongoing/future questions:



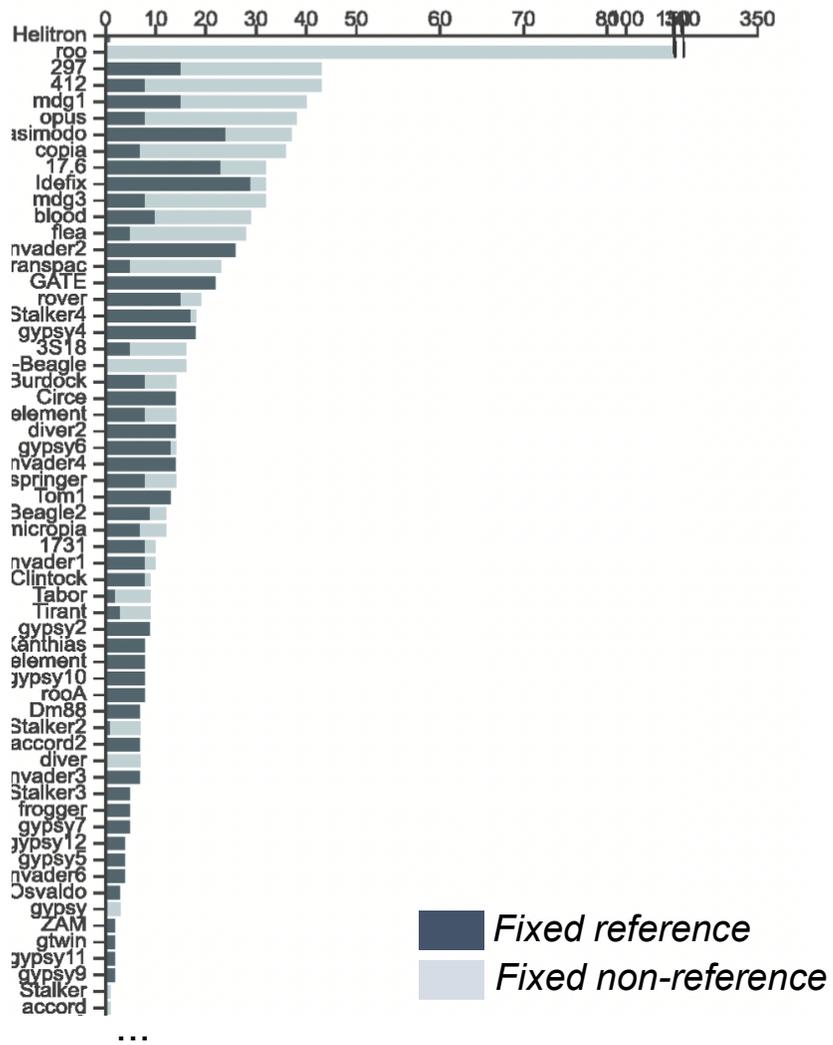
Precise TE landscape using long-read sequencing



Transposition upon aging?

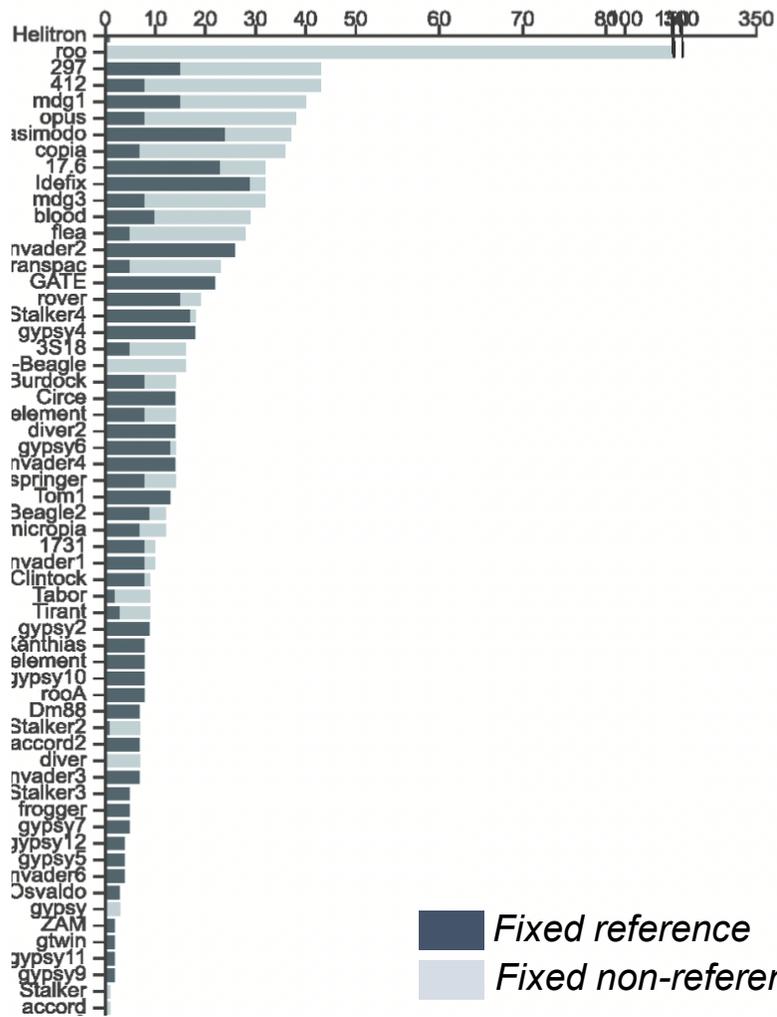
Selective somatic activity of LTR-retroelements

Fixed TE insertions per subfamily

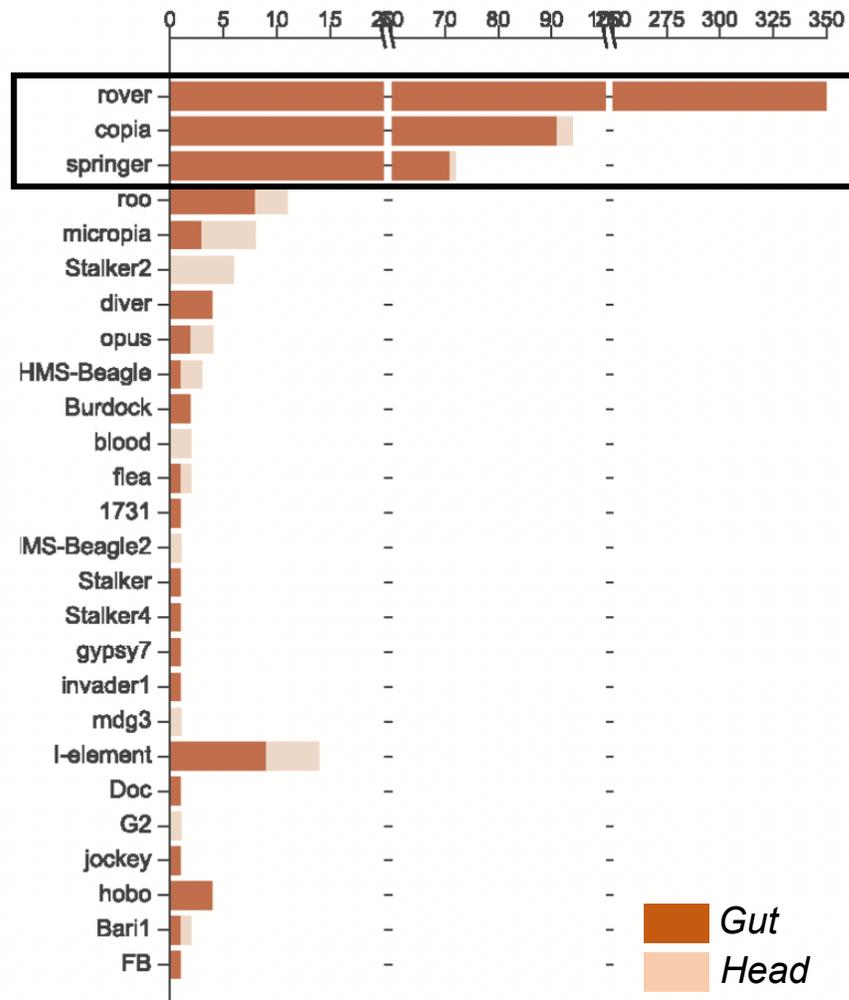


Selective somatic activity of LTR-retroelements

Fixed TE insertions per subfamily

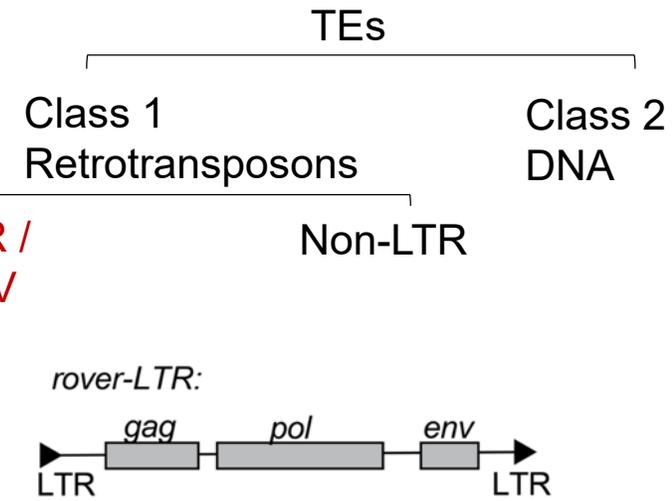


Somatic TE insertions per subfamily



Fixed reference
 Fixed non-reference

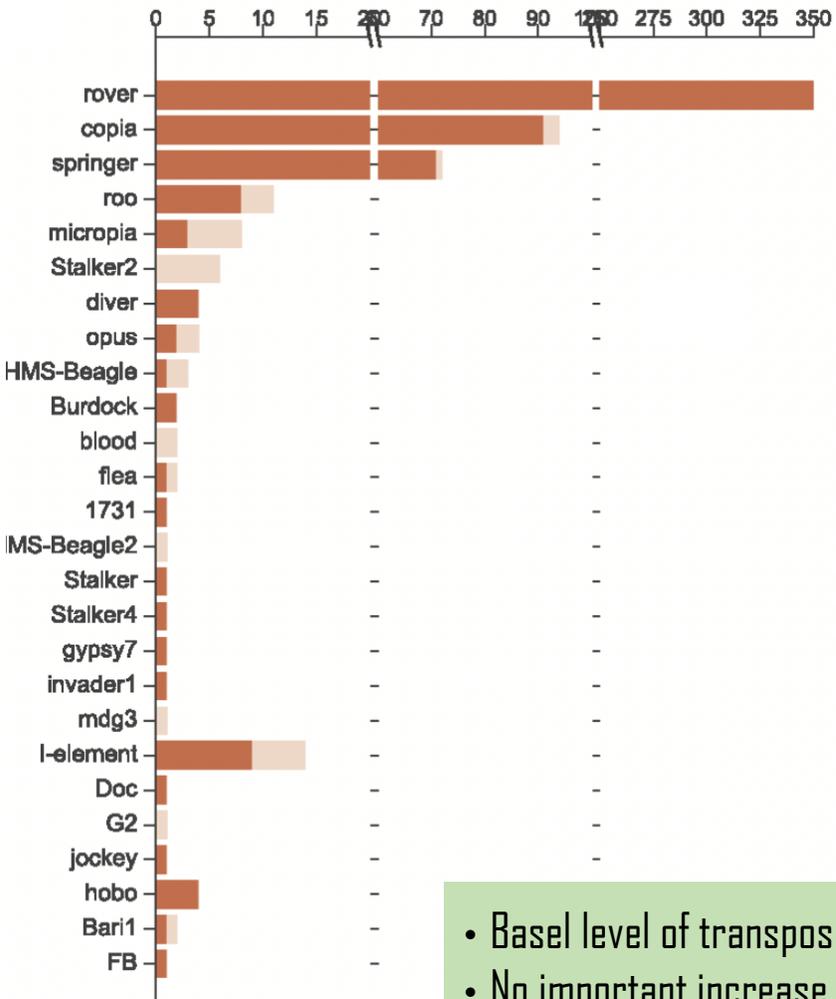
Gut
 Head



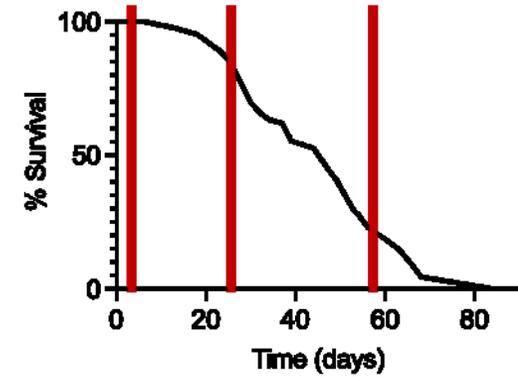
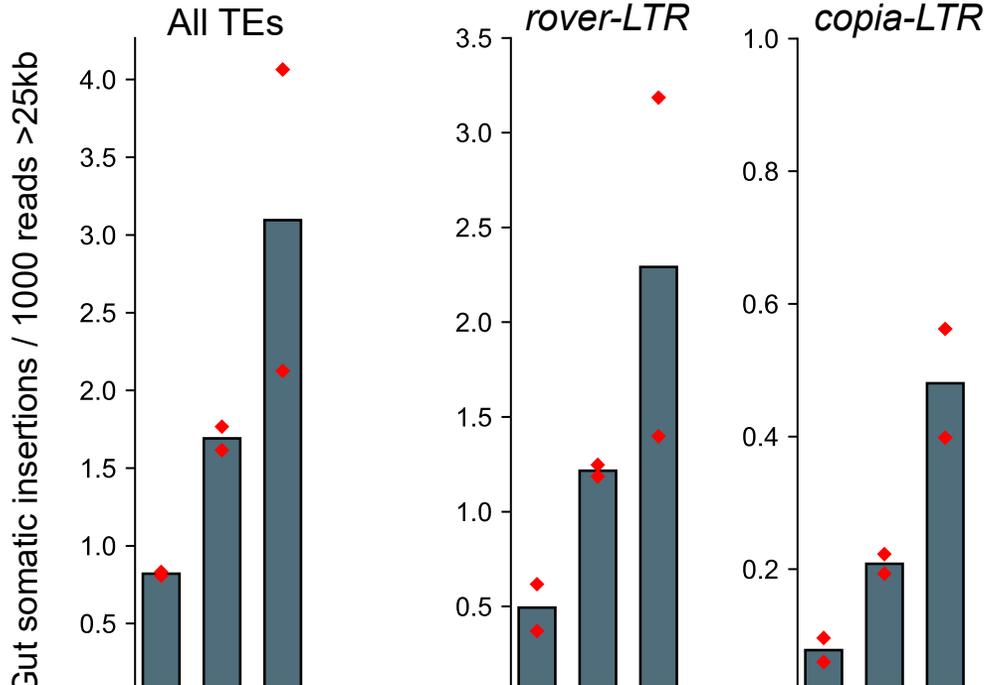
Is retrotransposon activity unleashed upon tissue aging?

Selective somatic activity of LTR-retroelements

Somatic TE insertions per subfamily



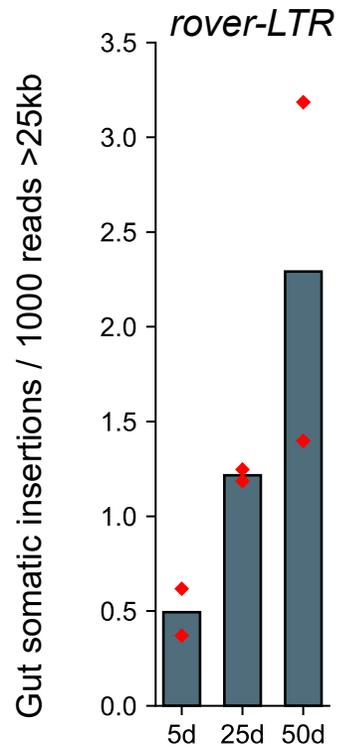
- Transposition upon aging?



- Basal level of transposition is limited to very few retroelements
- No important increase in TE mobility (rates or active families) in the aging gut (consistent with Yang et al. 2022 and Schneider et al. 2023)
- Impact on the tissue aging ?

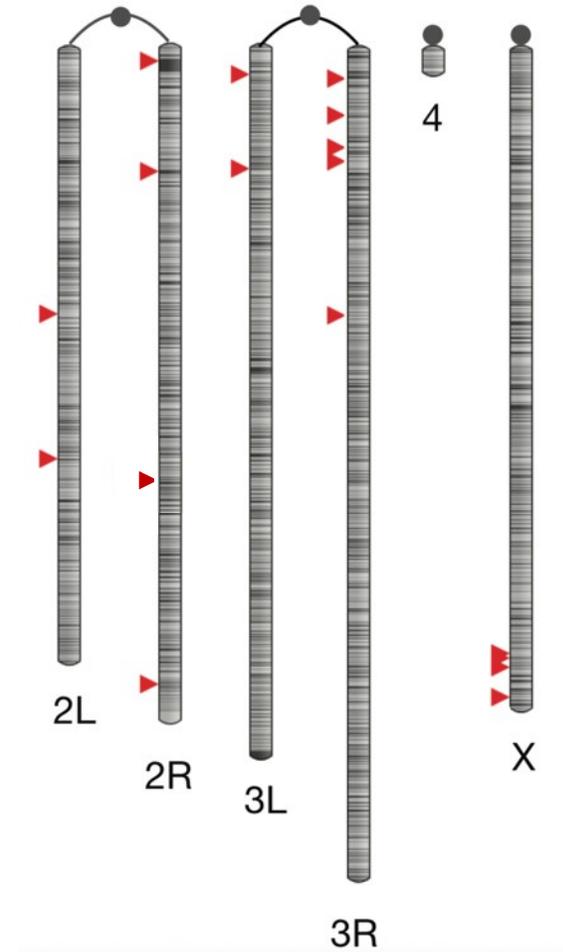
Why some retroelements are active (while not others)?

Why some retroelements are active (while not others)?



19 full-length *rover-LTR* elements in the genome:

How many active?



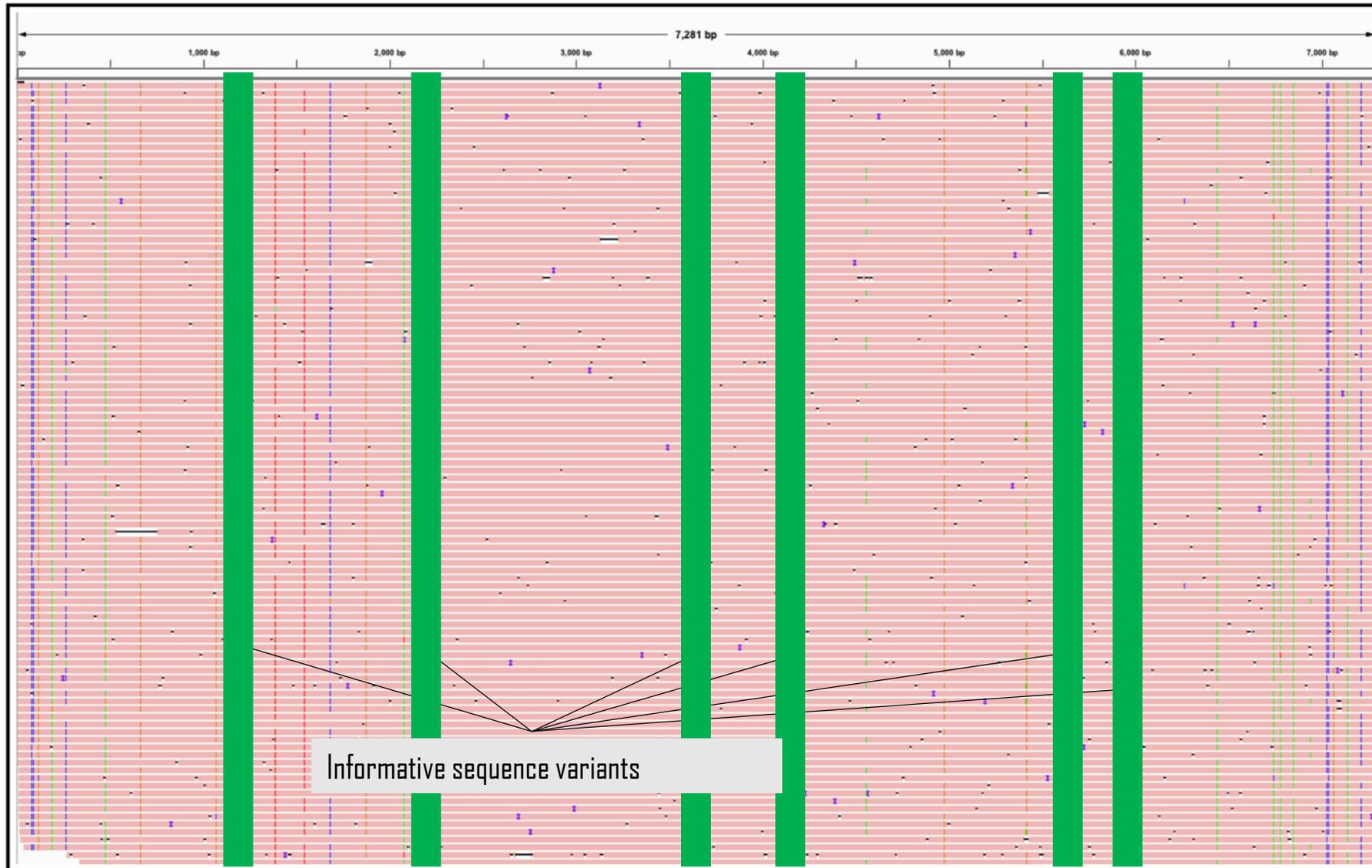
Identifying the active *rover*-LTR locus

- Recover and analyze sequences of somatic *rover* insertions from long-read data

canonical sequence:



somatic inserts:

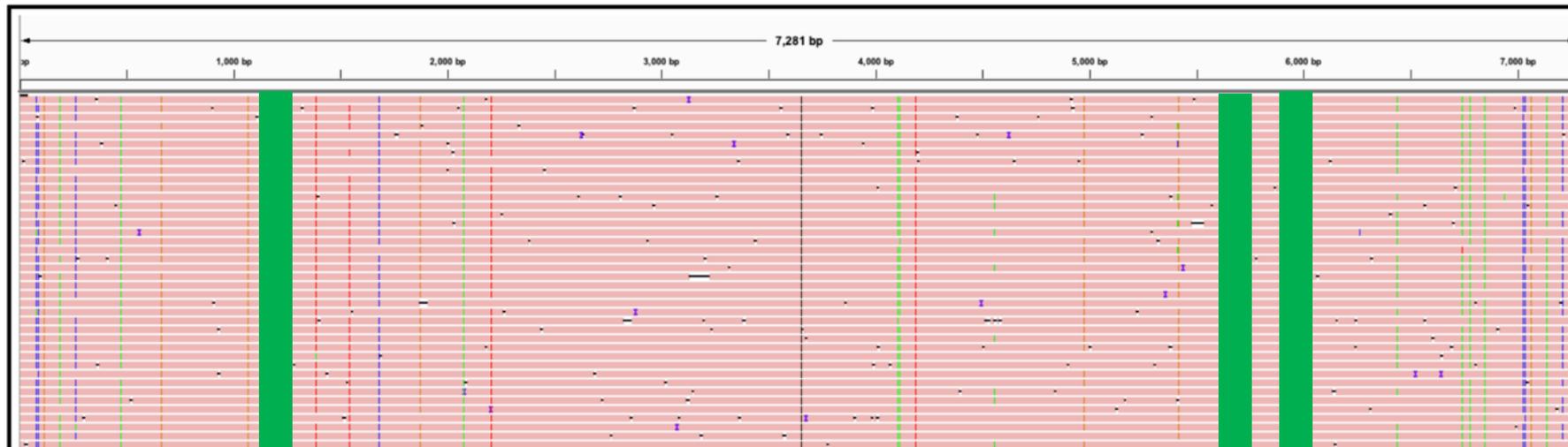


Identifying the active *rover*-LTR locus

- Compare somatic insertions with all fixed *rover* copies present in the genome (based on long-read data) Total = 19 copies

canonical sequence: 

somatic inserts:



candidate donors based on indel variants:

rover-2L:18M
rover-2R:14M



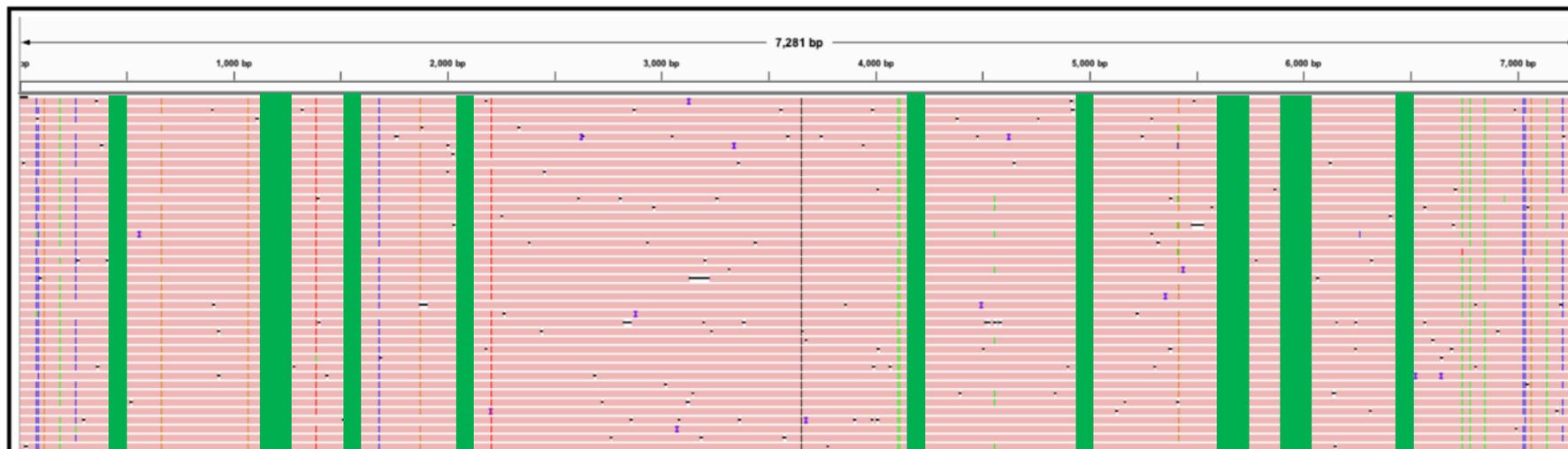
Identifying the active *rover*-LTR locus

- Compare somatic insertions with all fixed *rover* copies present in the genome (based on long-read data) Total = 19 copies

canonical sequence:



somatic inserts:



candidate donors based on indel variants:

rover-2L:18M
rover-2R:14M



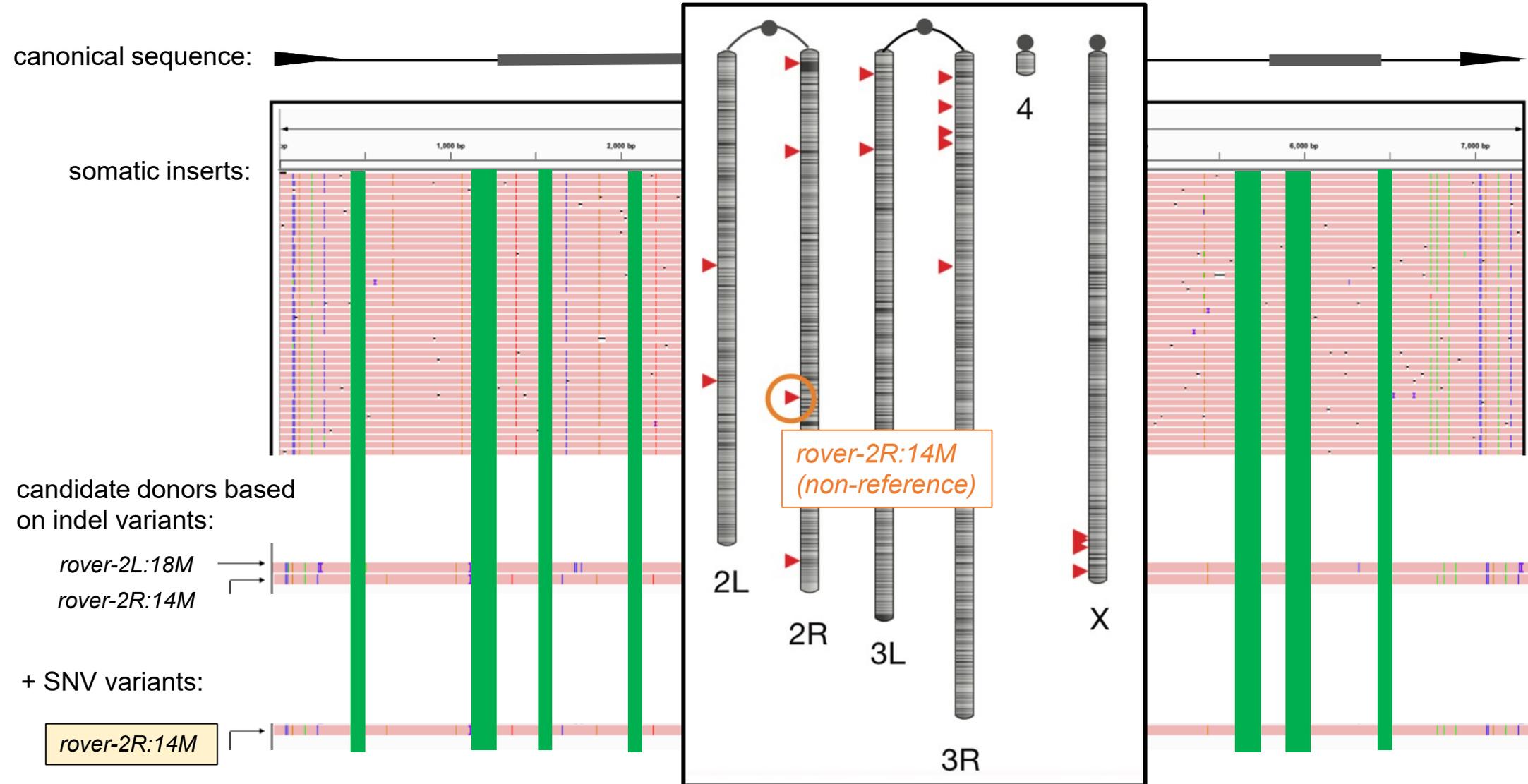
+ SNV variants:

rover-2R:14M

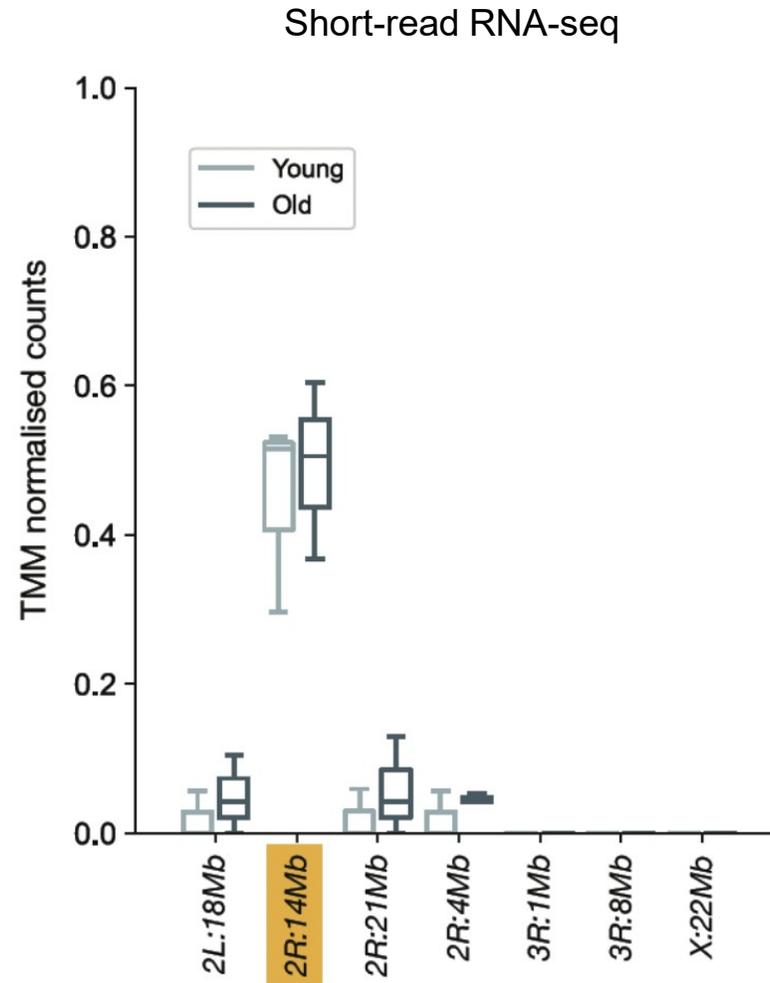
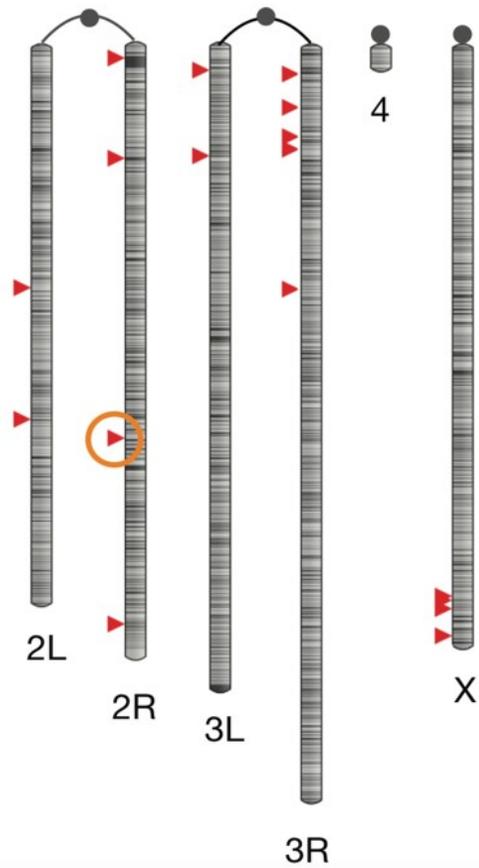


Identifying the active *rover*-LTR locus

- Compare somatic insertions with all fixed *rover* copies present in the genome (based on long-read data) Total = 19 copies



Transcriptome analysis supports expression of the active copy

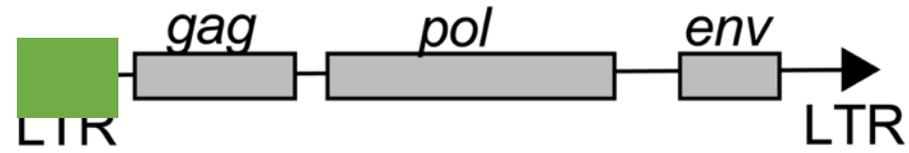


- Mobility of the *rover-LTR* subfamily is restricted to one genomic locus
- The first example of a "hot" LTR-retrotransposon locus active in a somatic tissue

Why ?

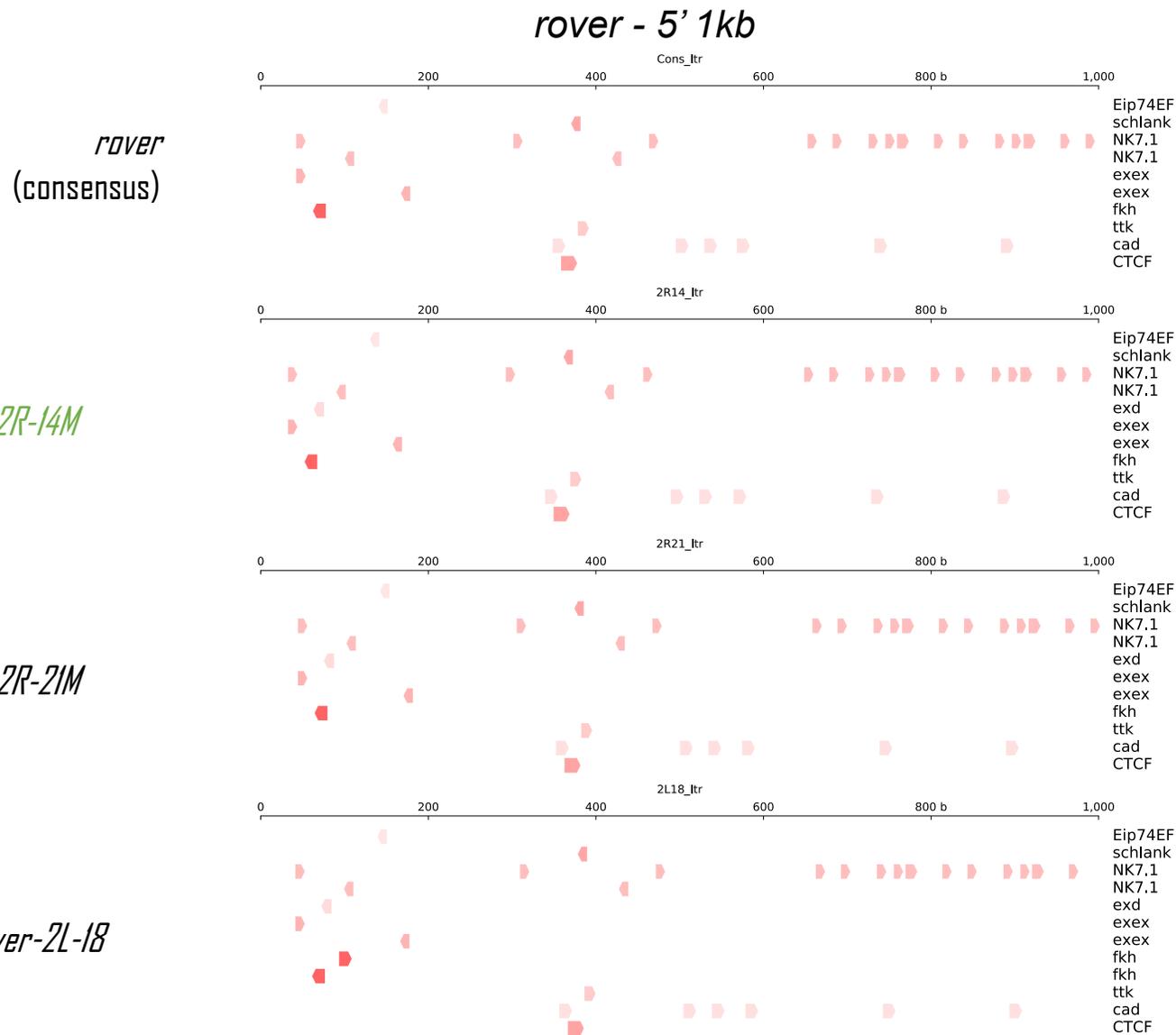
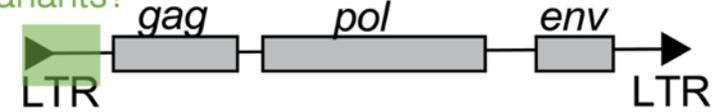
rover2R/4M 'hot' locus has little sequence variation in the regulatory region

Sequence variants?



rover 2R14M 'hot' locus has little sequence variation in the regulatory region

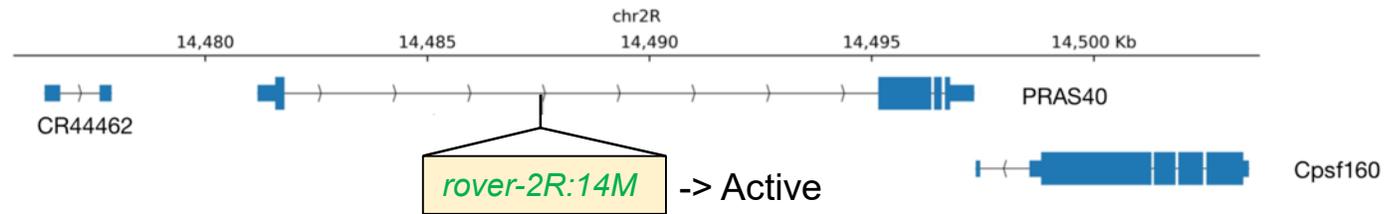
Sequence variants?



-> Activity not (only) due to TE sequence features but "local environment" ?

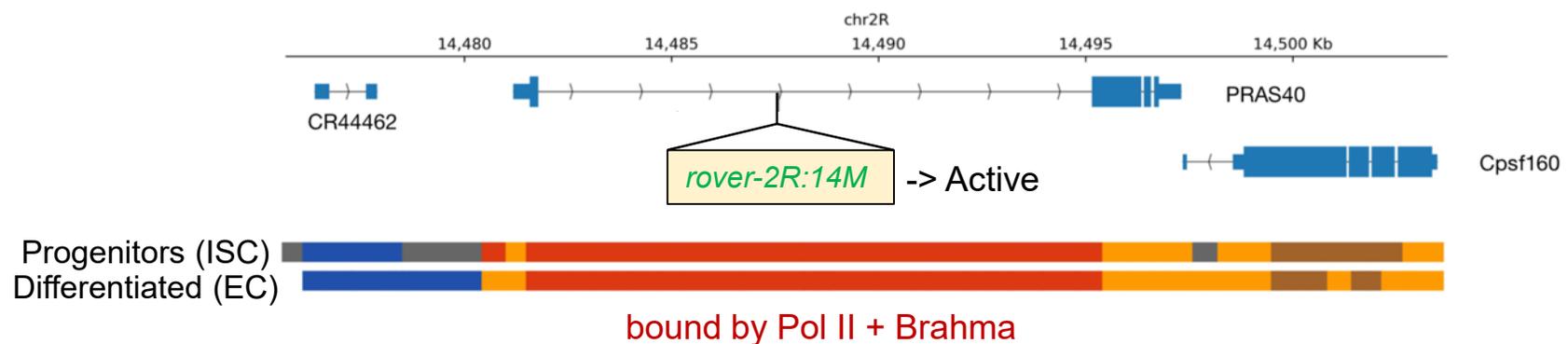
The active *rover* locus found in permissive chromatin

- Insertion in permissive chromatin likely allows expression



The active *rover* locus found in permissive chromatin

- Insertion in permissive chromatin likely allows expression



Cell type-specific
chromatin states:

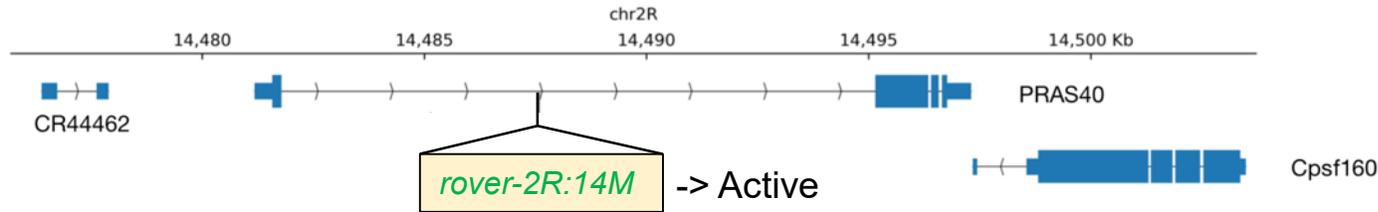
● ● Active

● ● ● Silent

(Gervais et al, 2019; Josserand et al, 2023)

The active *rover* locus found in permissive chromatin

- Insertion in permissive chromatin likely allows expression

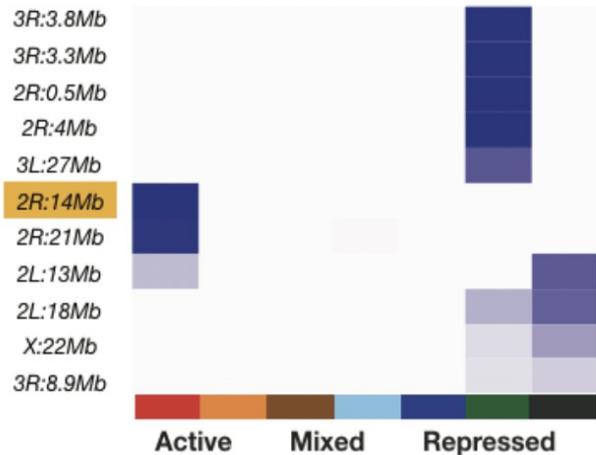


Cell type-specific chromatin states:

● Active
● Silent

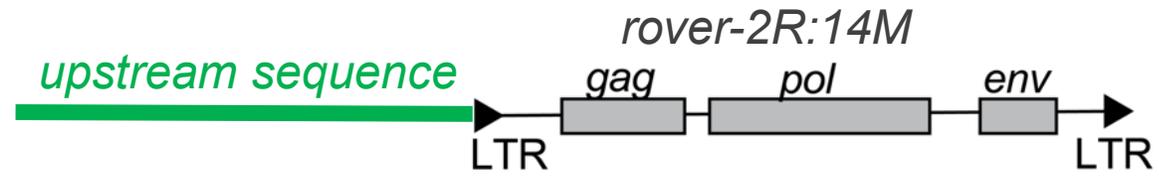
(Gervais et al, 2019; Josserand et al, 2023)

Chromatin states, ISC

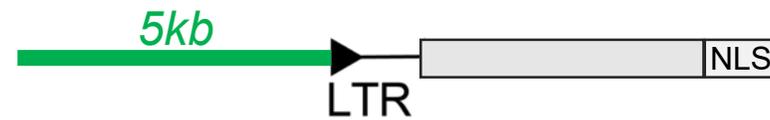


- Local genomic environment permits the expression (and the mobility) of the "hot" *rover2R/4M* locus

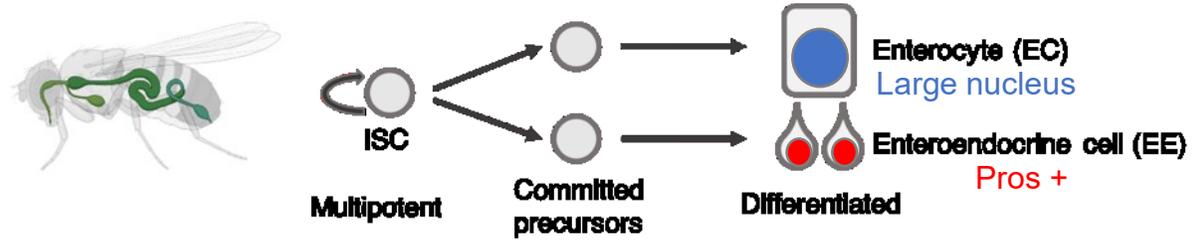
rover2R14M coopts upstream genomic sequence for gut expression



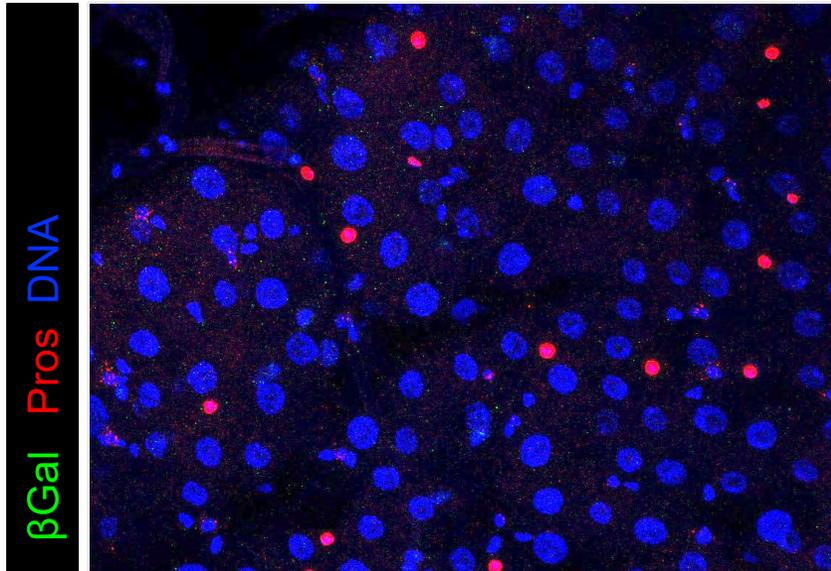
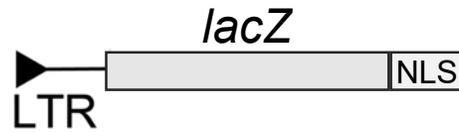
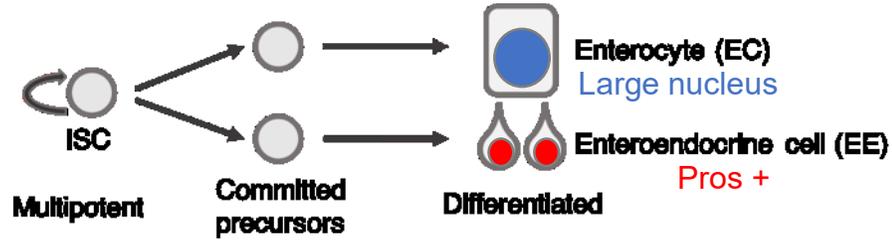
Expression reporters:



rover2R14M co-opts upstream genomic sequence for gut expression

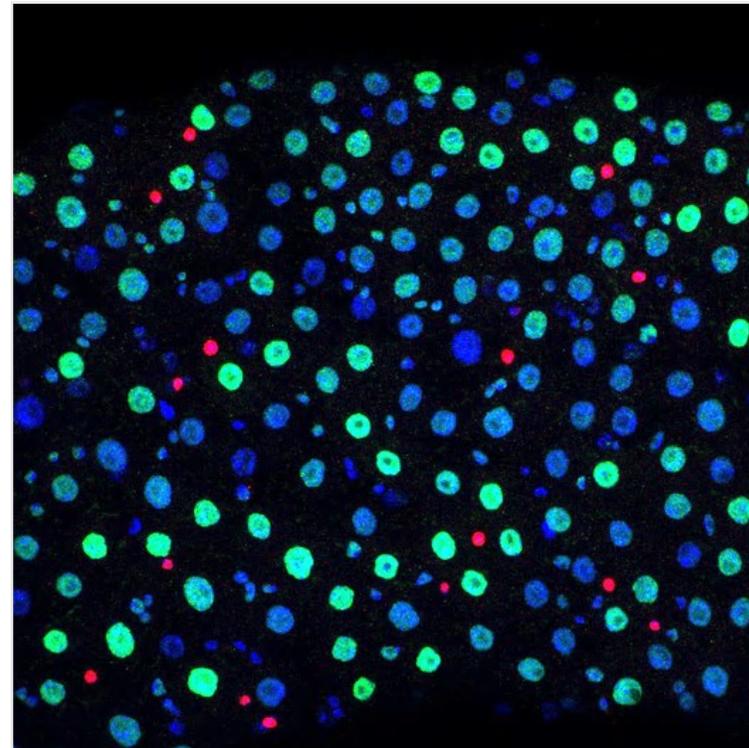


rover2R14M co-opts upstream genomic sequence for gut expression



βGal Pros DNA

As shown before, with functional small RNA silencing pathways
(Wang et al, 2022)



Do tissue-specific TFs drive *rover2R:14M* expression ?



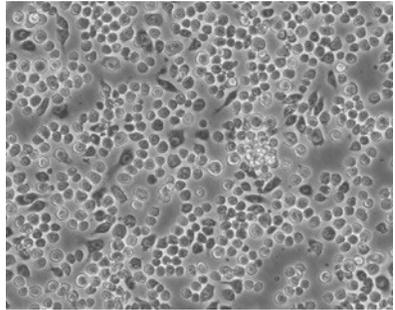
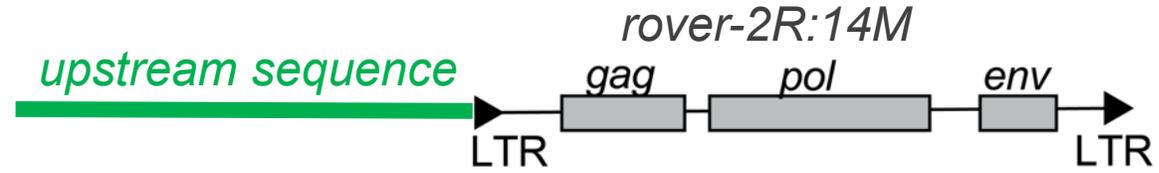
Mapped enhancers
(whole organism, SEA – Chen et al, 2019)

ATAC-seq picks
(gut progenitor cells- Tauc et al, 2021)

ChIP-seq picks
(different tissues/stages - ModENCODE)

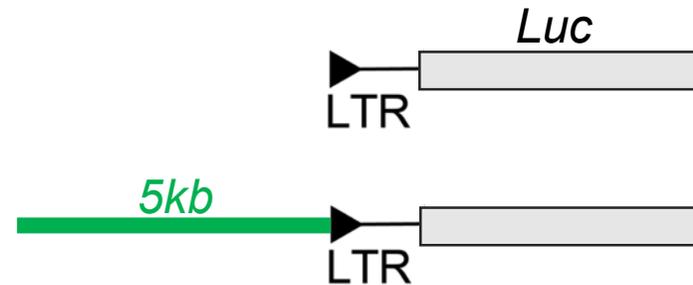


Do tissue-specific TFs drive *rover2R:14M* expression ?



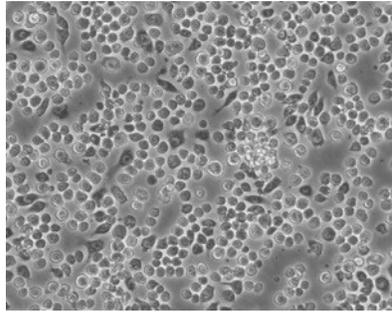
S2 cells

Luciferase expression reporters:



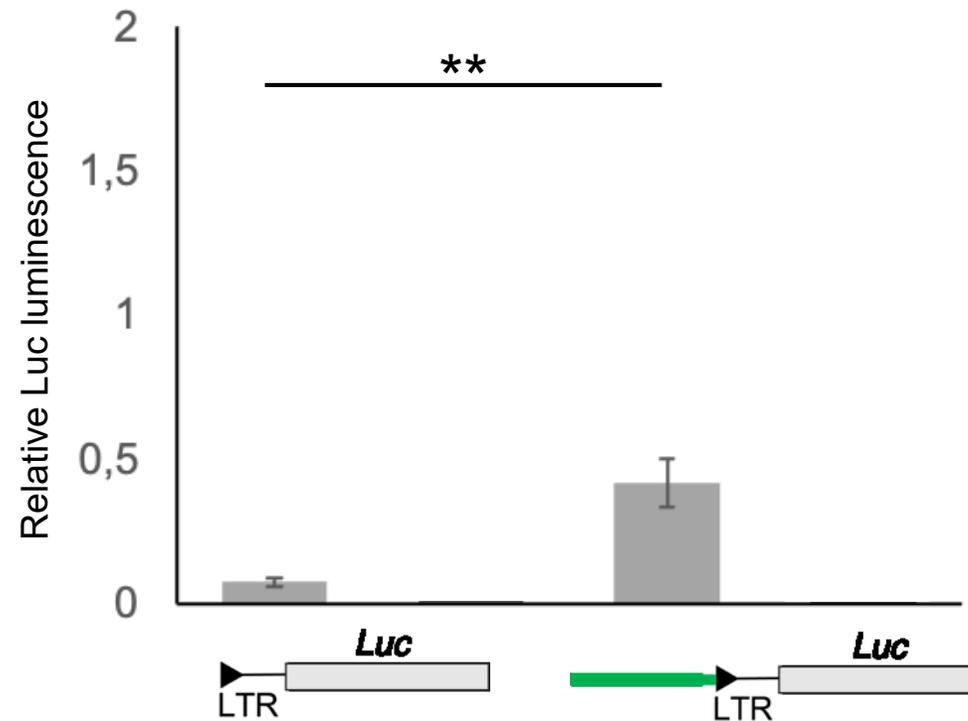
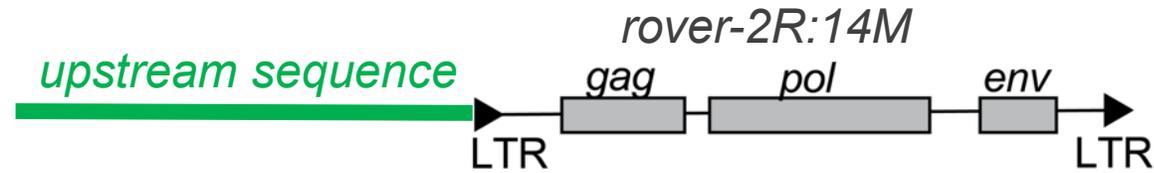
+ / - candidate TFs

Escargot may drive *rover2R:14M* expression via the upstream sequence

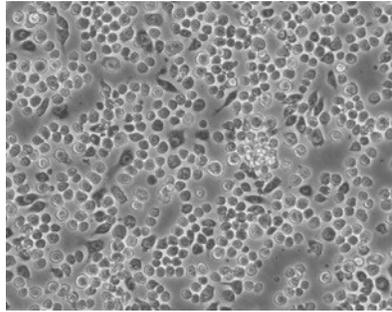


S2 cells

Luciferase expression reporters:

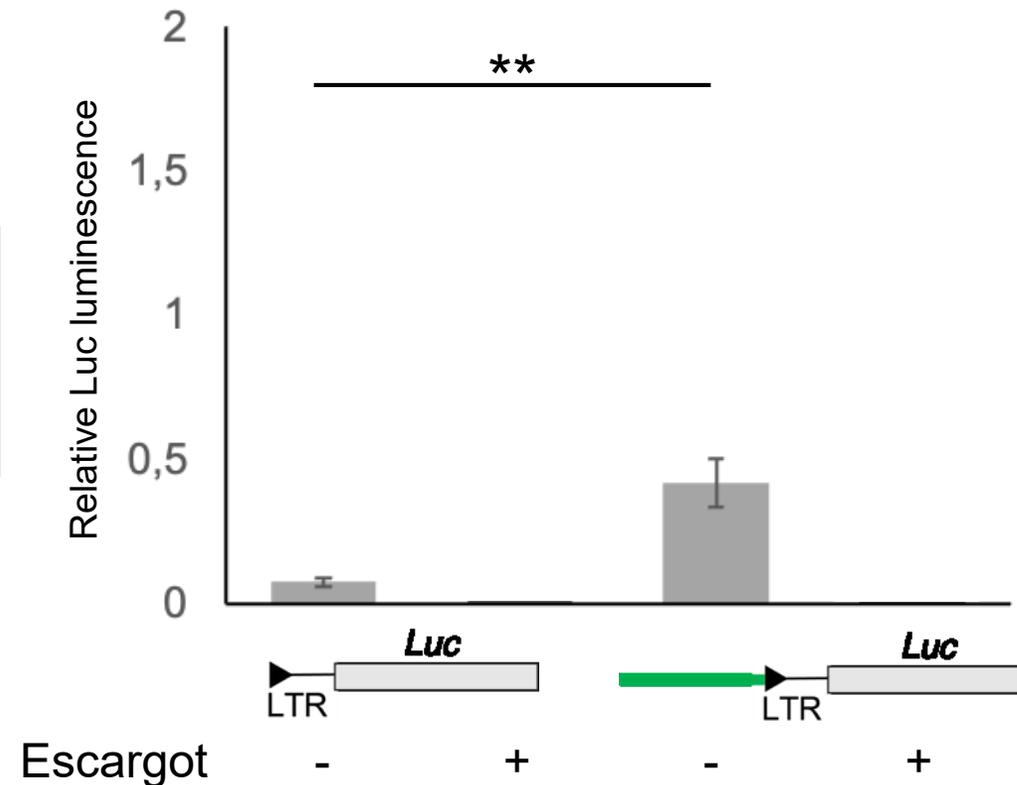
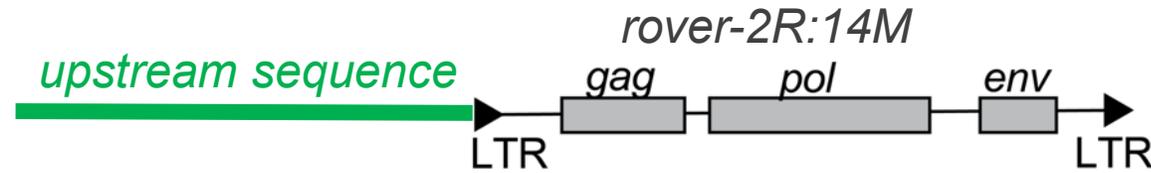


Escargot may drive *rover2R:14M* expression via the upstream sequence



S2 cells

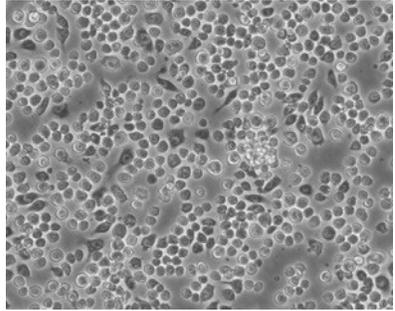
Luciferase expression reporters:



Escargot (*esg*):

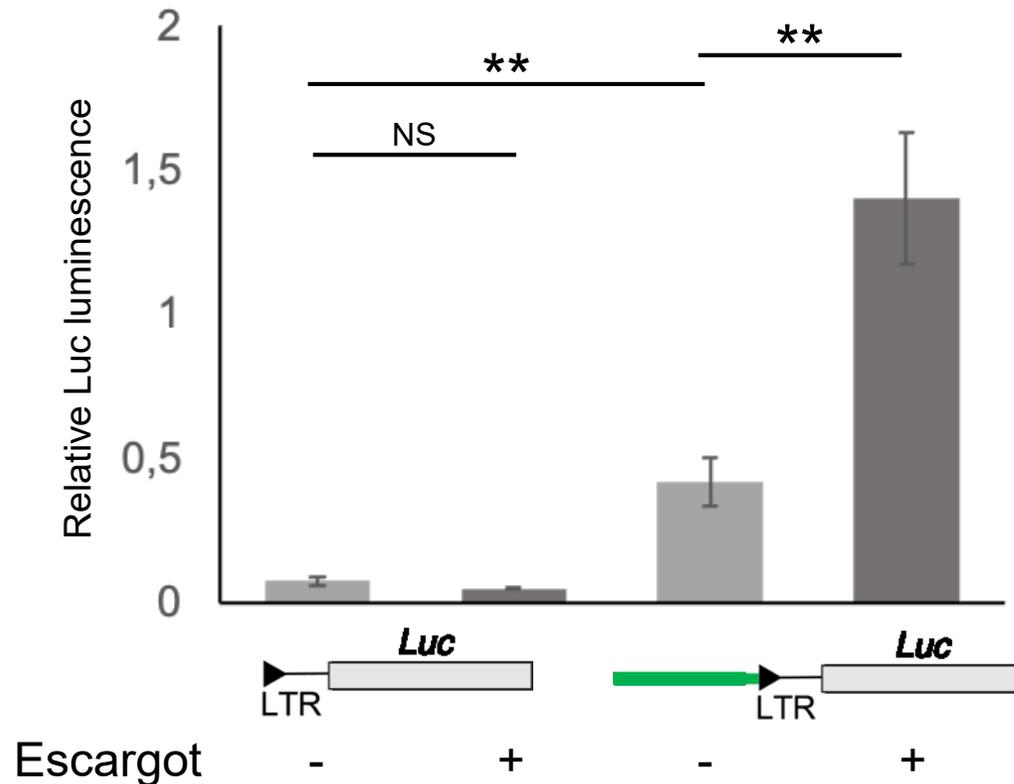
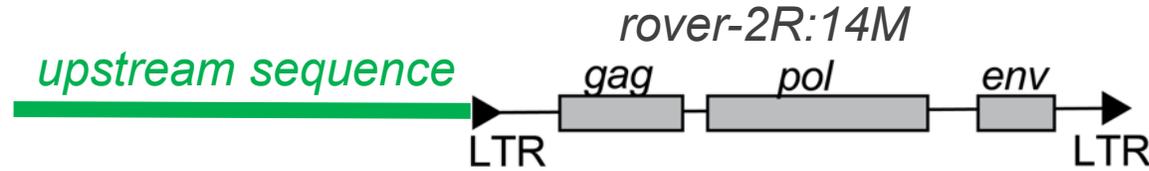
- Snail-type TF
- Enriched in gut progenitors
- Stem cell maintenance

Escargot may drive *rover2R:14M* expression via the upstream sequence



S2 cells

Luciferase expression reporters:



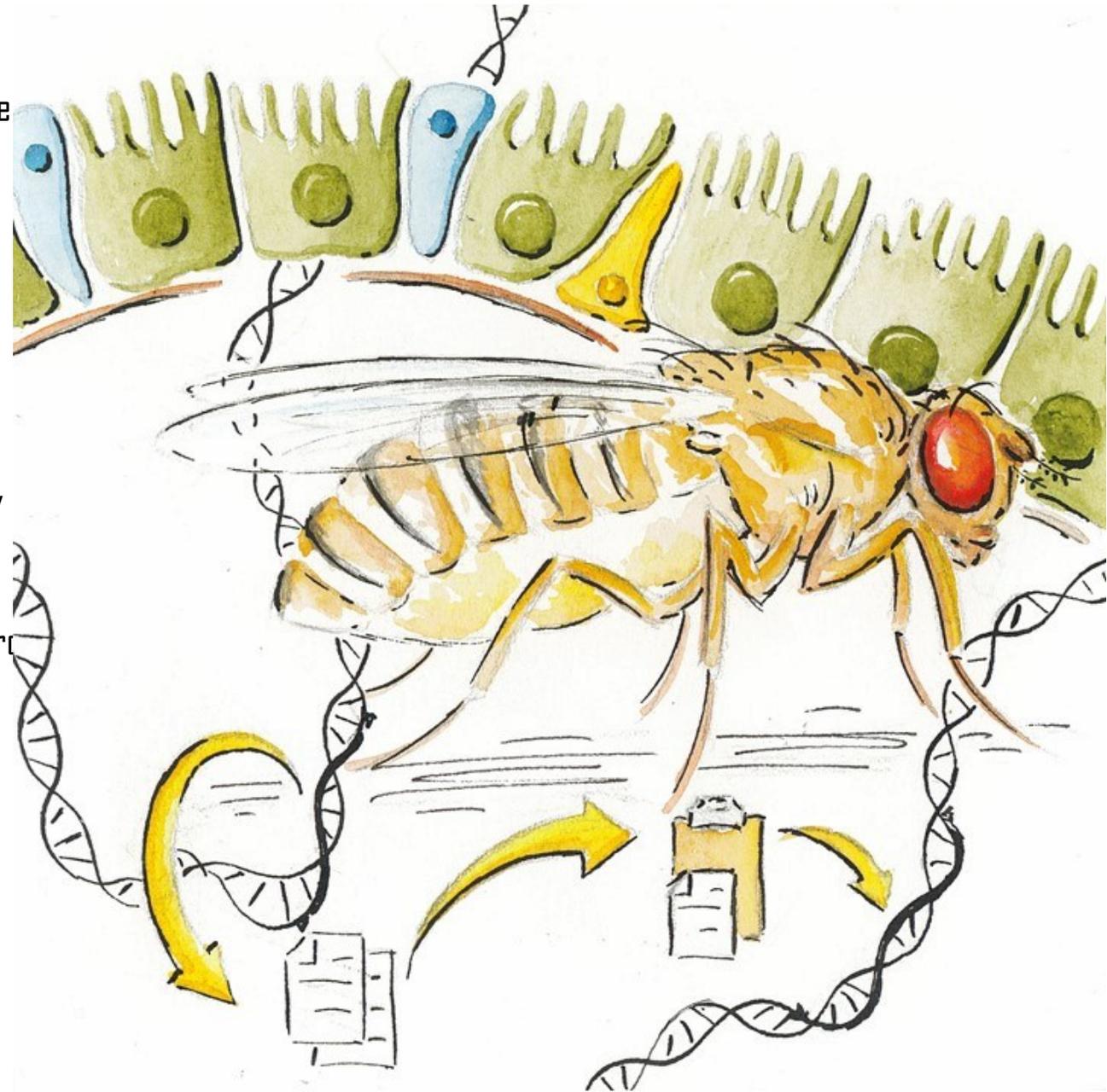
Escargot (*esg*):

- Snail-type TF
- Enriched in gut progenitors
- Stem cell maintenance

Other candidate TFs ?
(ongoing)

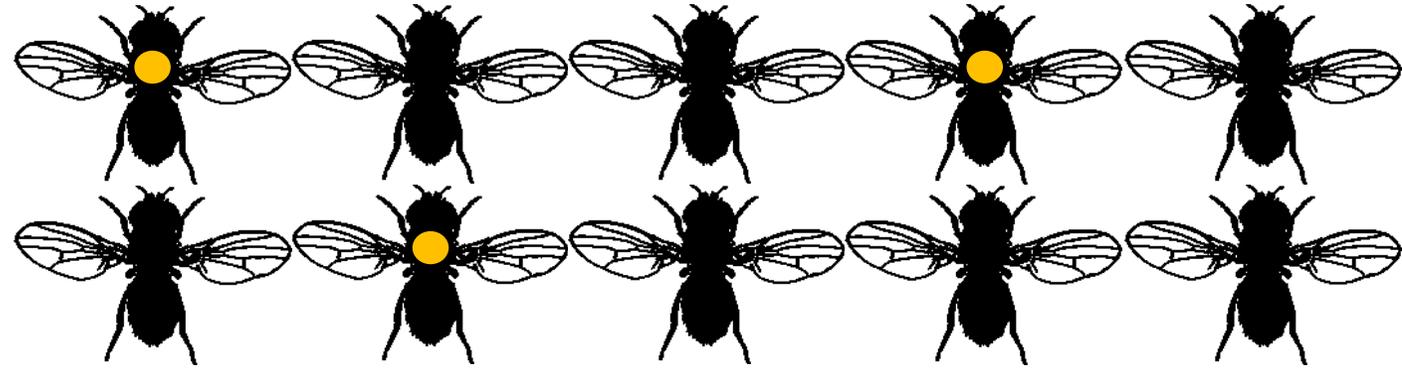
Conclusion 2

- Long-read DNT sequencing allowed full-length analysis of fixed mobile throughout different adult ages
- No significant increase in retrotransposon mobility in the aging gut
- Basel level of transposition is limited to very few retroelements
- We identify a “hot” somatically active LTR-retroelement : *rover-2R:14M*
- *rover2R14M* co-opts upstream genomic sequence for gut expression through



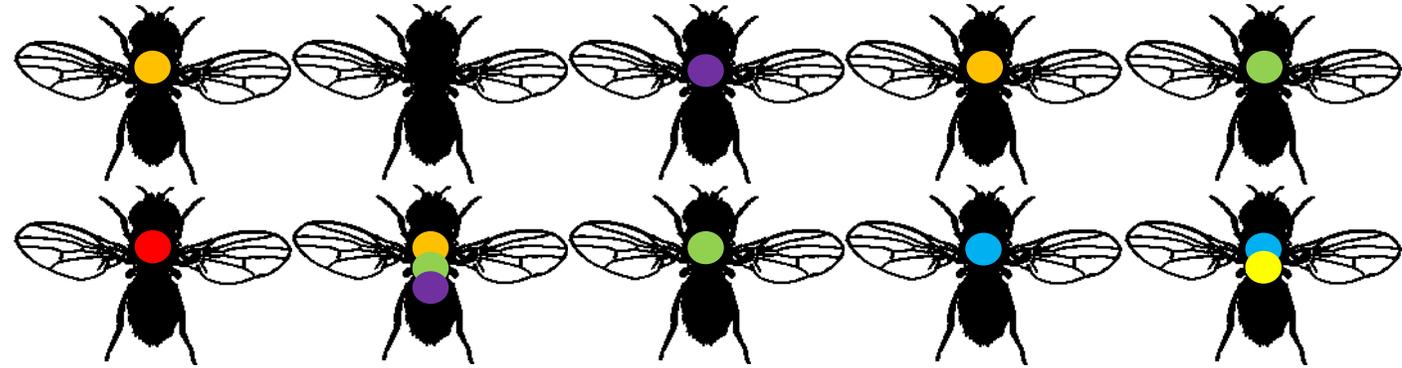
Impact ?

- **rover-2R:14M** is a non-reference polymorphic TE locus



Impact ?

- **rover-2R:14M** is a non-reference polymorphic TE locus
- TE polymorphism is large in populations

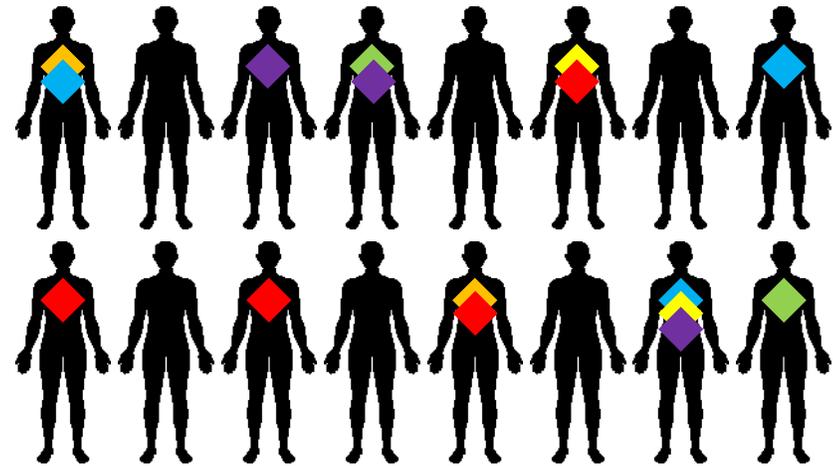


How does this variation contribute to somatic TE activity ?

... and phenotypic differences ?

Working on TEs ?

Mind the genetic backgrounds !



Around 1500 estimated private insertions / human genome

(C. Beck lab, unpublished communication)

Thank You !



Allison Bardin
and the team

Past members:
Natalia Rubanova
Marius van den Beek
Nick Riddiford

Bioinformatics:
Nicolas Servant

Fly facility:
Bénédicte Leveille Nizerolle
Sebastian Bordeaux Montrieux

Sequencing Facility:
Sonia

Louis Barolle
Fabienne Chalvet
Sophie Netter
Mickael Poidevin
Darshika Singh
Maxence de Taffin de Tilques

Fly facility: Delphine Petit

NGS Sequencing Facility

Imaging Facility



StG 2023: gutTEimpact