BovReg
Understanding cattle genomes

GWAS reveals determinants of mobilization rate and dynamics of an active endogenous retrovirus of cattle

Lijing Tang, Benjamin Swedlund, Sébastien Dupont, Chad Harland, Gabriel Costa Monteiro Moreira, Keith Durkin, Maria Artesi, Eric Mullaart, Arnaud Sartelet, Latifa Karim, Wouter Coppieters, Michel Georges, Carole Charlier (ULIEGE)

Nature Communications, in press

BovReg Final Conference - Brussels

(14-15 February 2024)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668

Source of genetic variation: germline *de novo* mutations

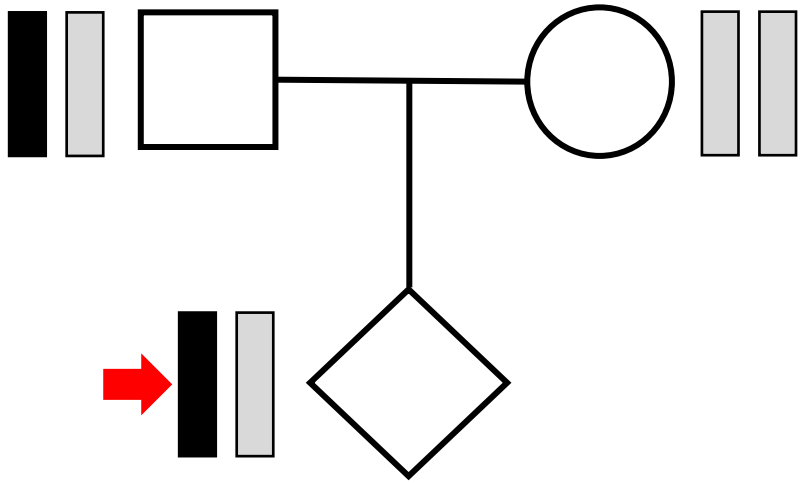
Polymorphisms → Darwinian selection →

LOSS from random drift

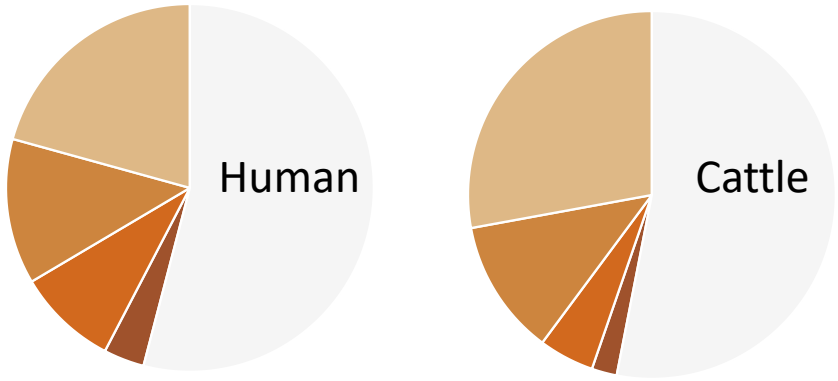
GAIN from *de novo* mutations

→ *de novo* mutations (➡)

Polymorphisms
SNP, INDEL, CNV, ...
mobilization of transposons



The mammalian transposon landscape: many dead, few still living



Genome TE contents

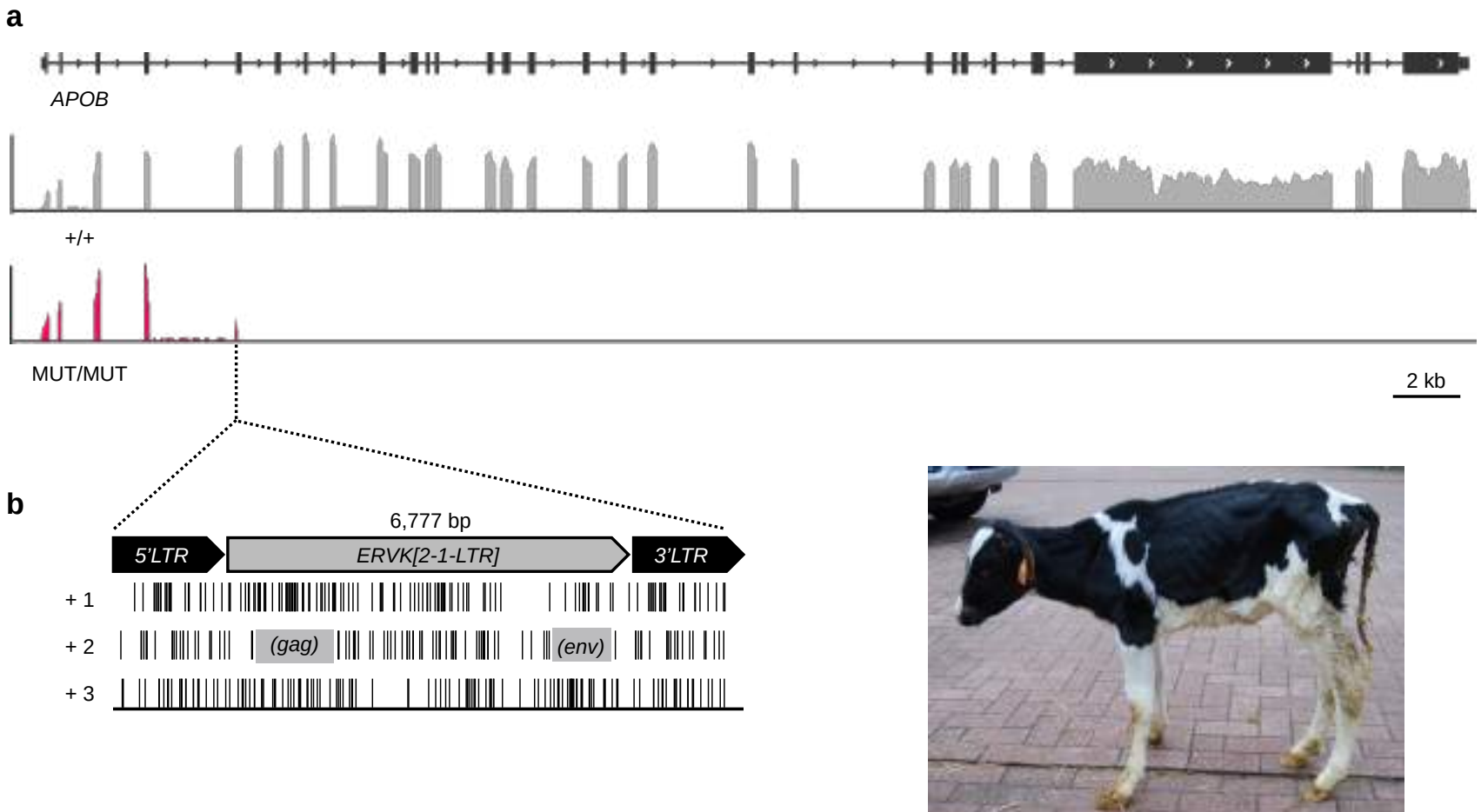
- LINE: Long Interspersed Nuclear Element
- SINE: Short Interspersed Nuclear Element
- ERV: Endogenous Retrovirus
- DNA: DNA transposon
- Non - TE

TE: Transposable Elements

Barbara McClintock

LINE1	SINE	ERV	DNA
✅	✅		
✅	✅	?	

ERVK elements might still be mobile in cattle *APOB* mutation causing cholesterol deficiency

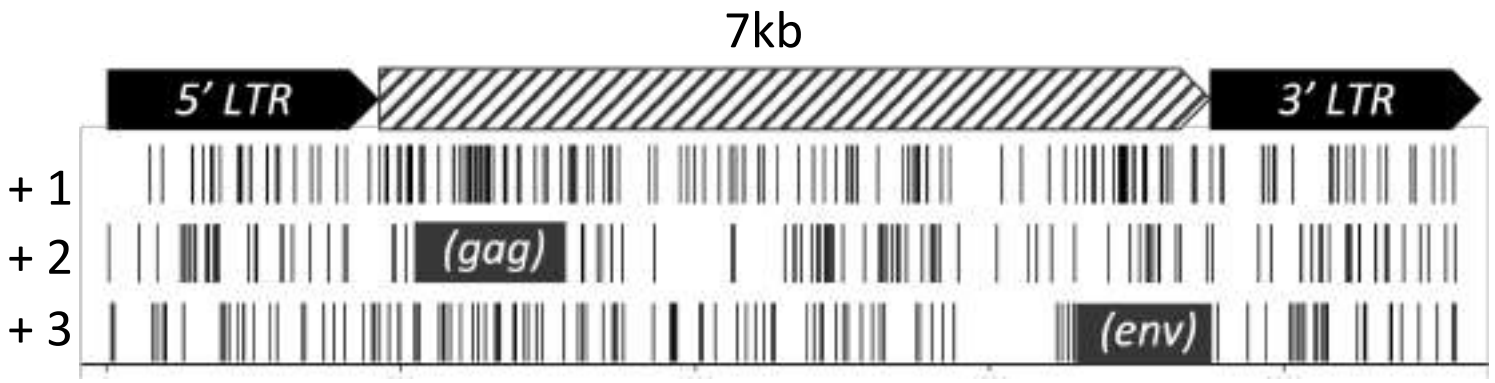


OMIA:001965-9913

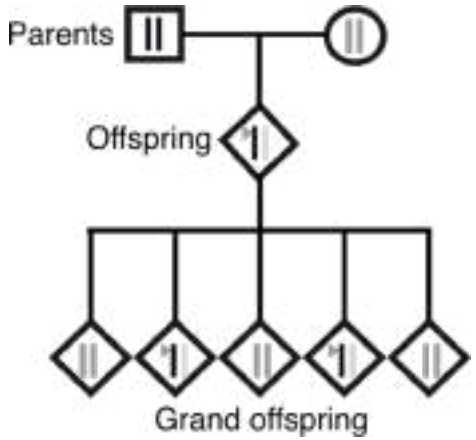
2015

ERVK elements are still active in the bovine paternal and maternal germline (Damona)

Insertion site (bp)	Feature	Orientation	Gene	Germline	G-off
chrX:35057434-35057439	Exonic	Sense	*GABRQ*	Dam3	3/5
chr2:38898428-38898436	Intronic	Sense	*CYTIP*	Sire1	2/5
chr5:68225832-68225837	Intronic	Sense	*CHST11*	Sire1*	2/5
chr18:9006859-9006864	Intergenic	/	/	Sire1*	3/5
chr19:4744520-4744529	Intergenic	/	/	Sire3	2/5



N = 127



2 X 127 gametes

5 *de novo* ERVK insertions

3 S_1 / 1 S_2 / 0….0 other Sires

1 D_1 / 0….0 other Dams

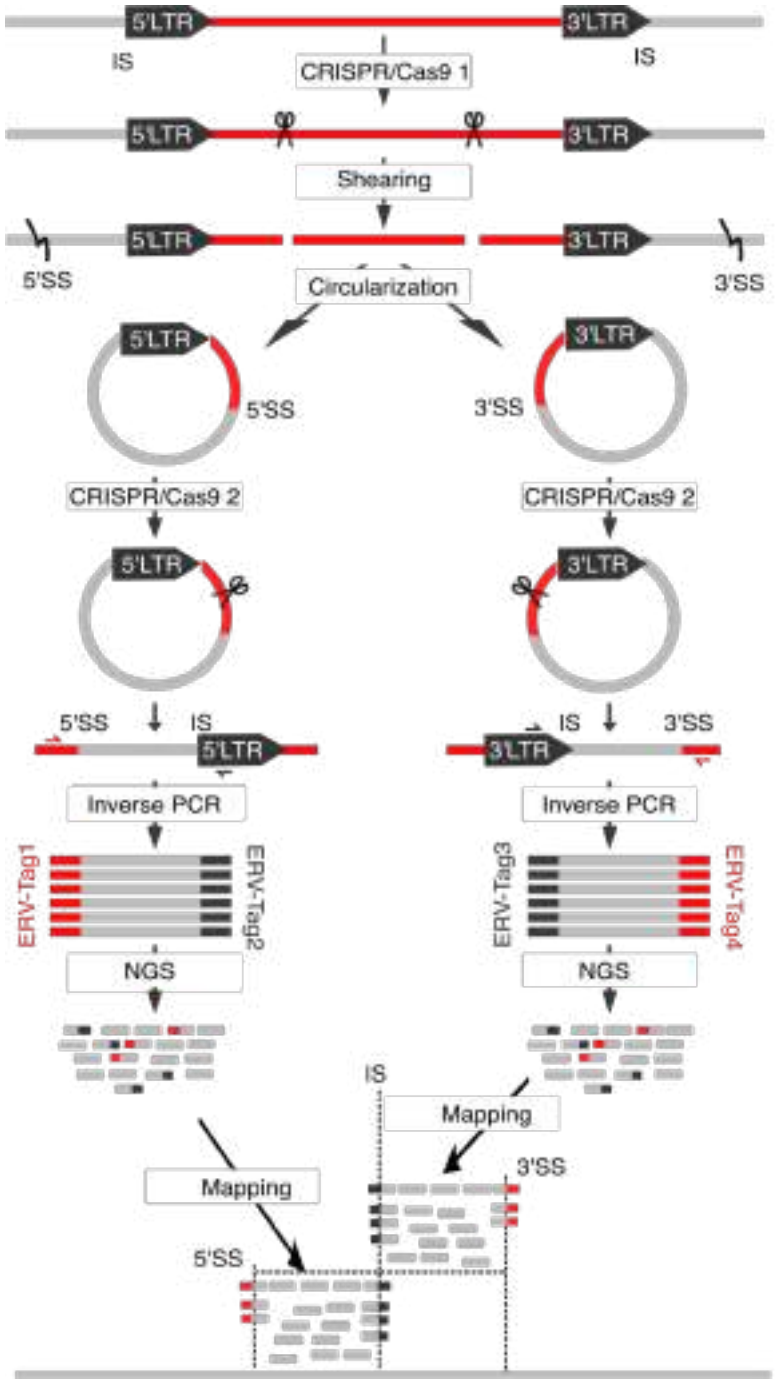
De novo rate: 5/254 ~1/50

Damona pedigree
Holstein Friesian
753 WGS
5 *de novo* ERVK events
Variable ?

Scaling up needed…

Direct quantitative measurement ?
Belgian Blue breed

‘PCIP’

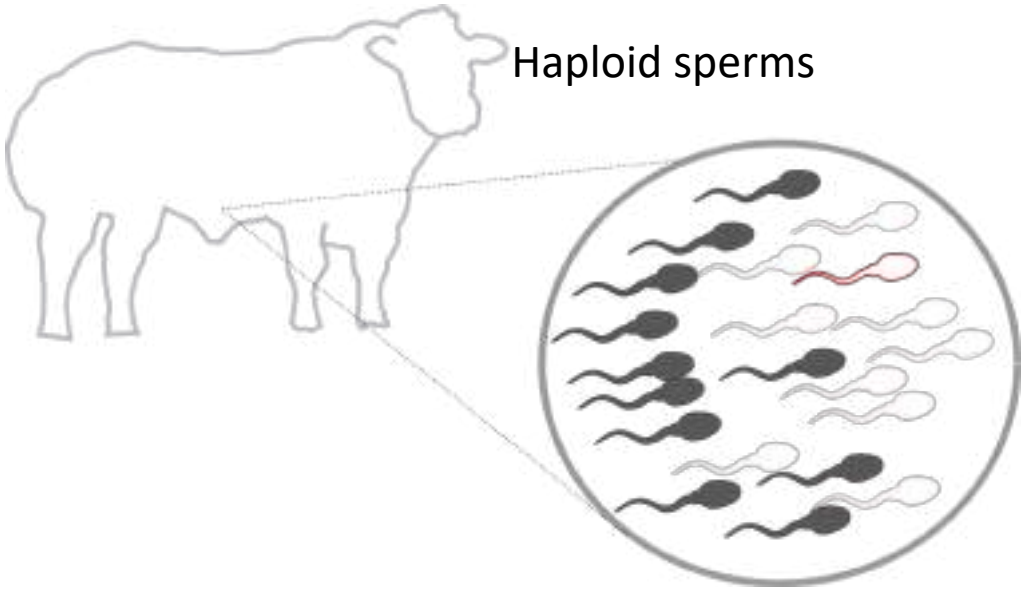


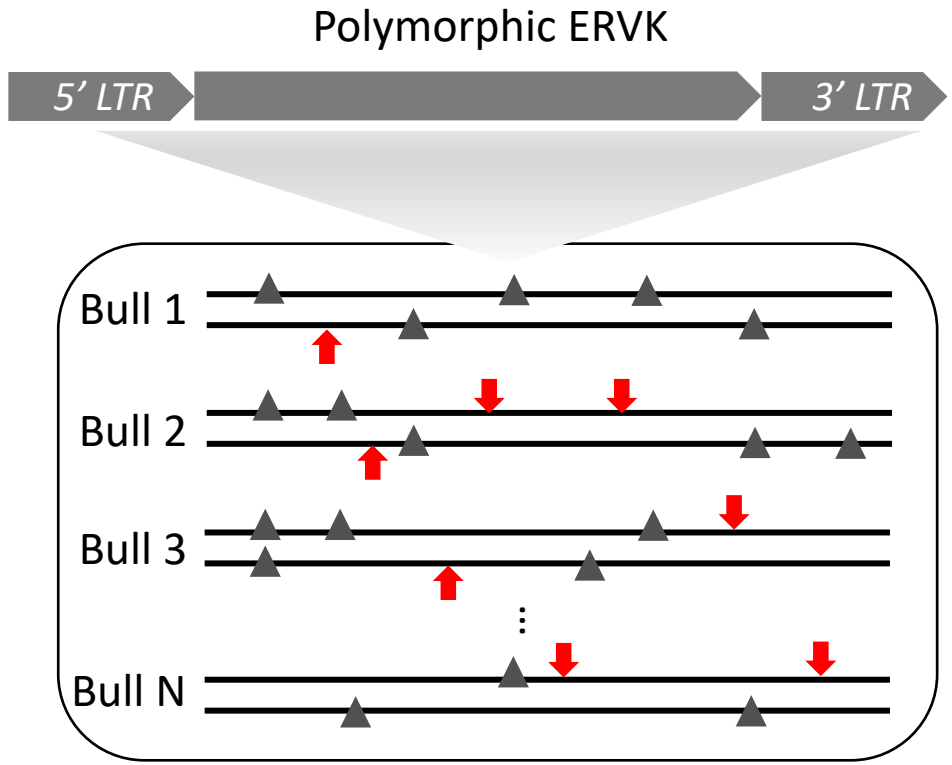
We developed of a method to directly measure ERVK mobilization rate in germline

- Could be applied to any class of active transposable elements
- Provides technical replicates by targeting both 5’ and 3’ end
- Can be rendered quantitative by using the polymorphic sites inherited from the parents as internal controls

Artesi et al., 2021

We developed of a method to directly measure ERVK mobilization rate in the germline





$$\textit{Transposition rate (dnTR)} = \frac{\textit{Number of detected de novo loci}}{\textit{Effectively captured haploid genomes}}$$

▲ Polymorphic ERVK (306)

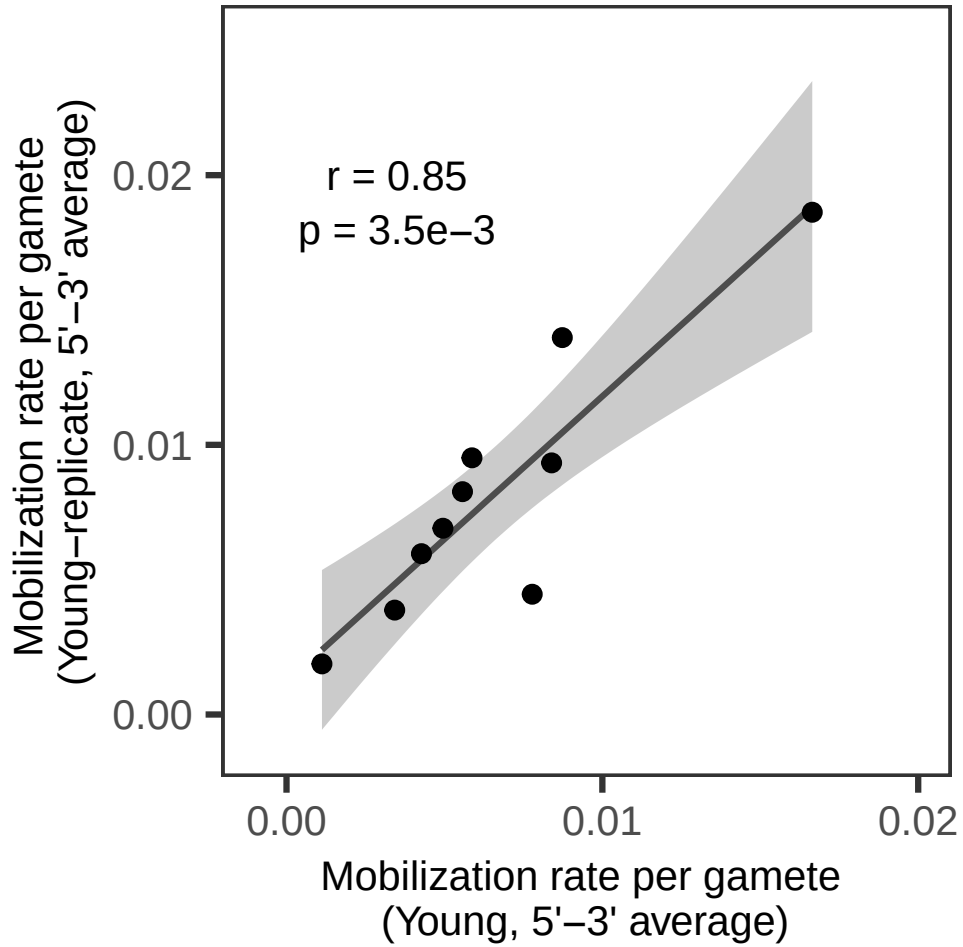
De novo ERVK

〓 Diploid genome

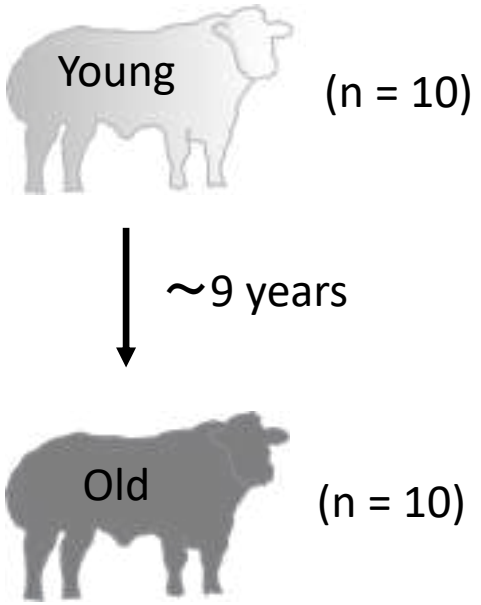
The mobilization rate of ERVK in the male germline is a highly repeatable phenotype

Young R1 (n = 10)

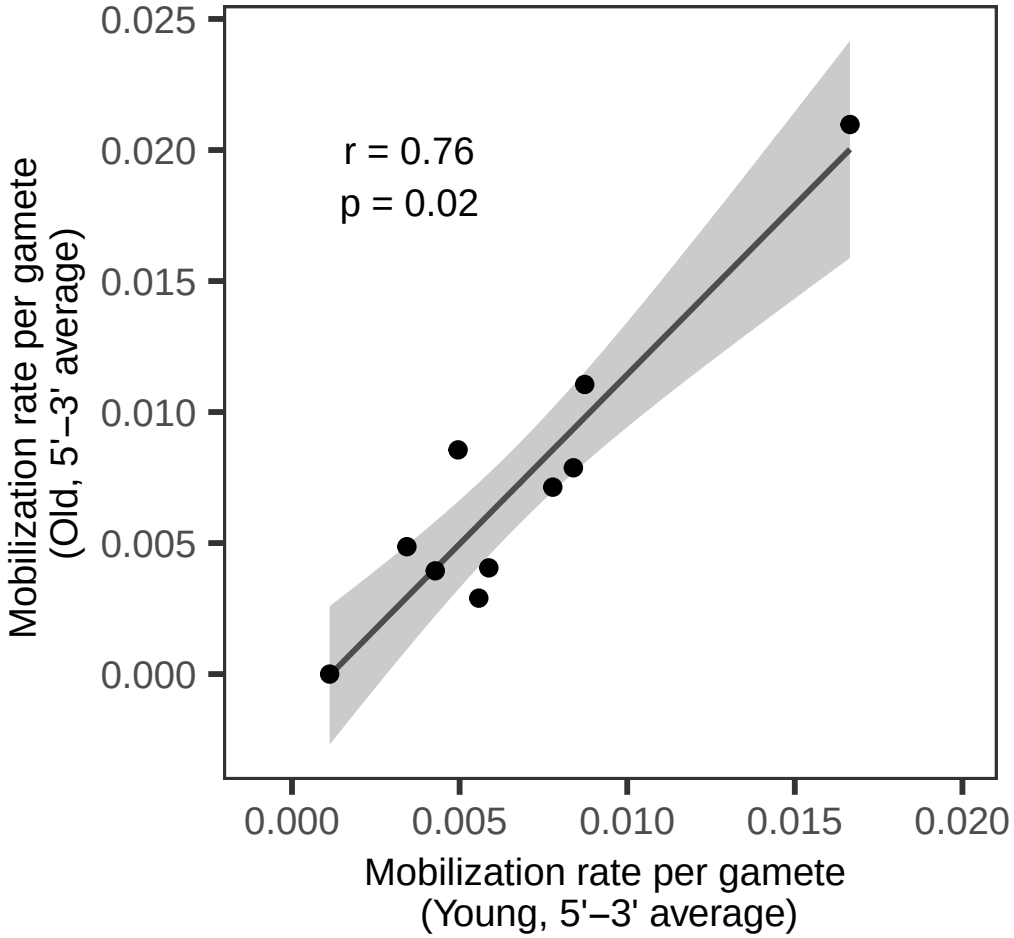
Young R2 (n = 10)



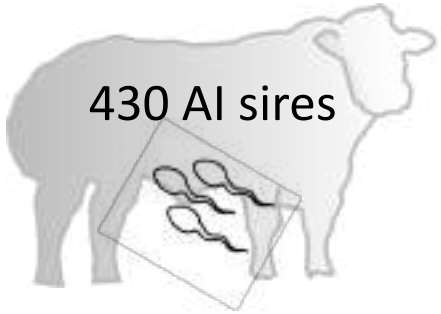
The ERVK mobilization rate in the male germline is stable throughout the lifetime



No evidence of an effect of age on mobilization rate

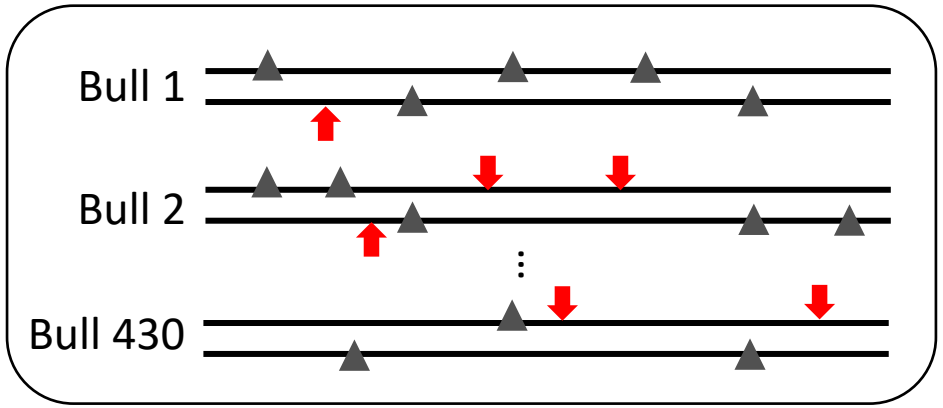


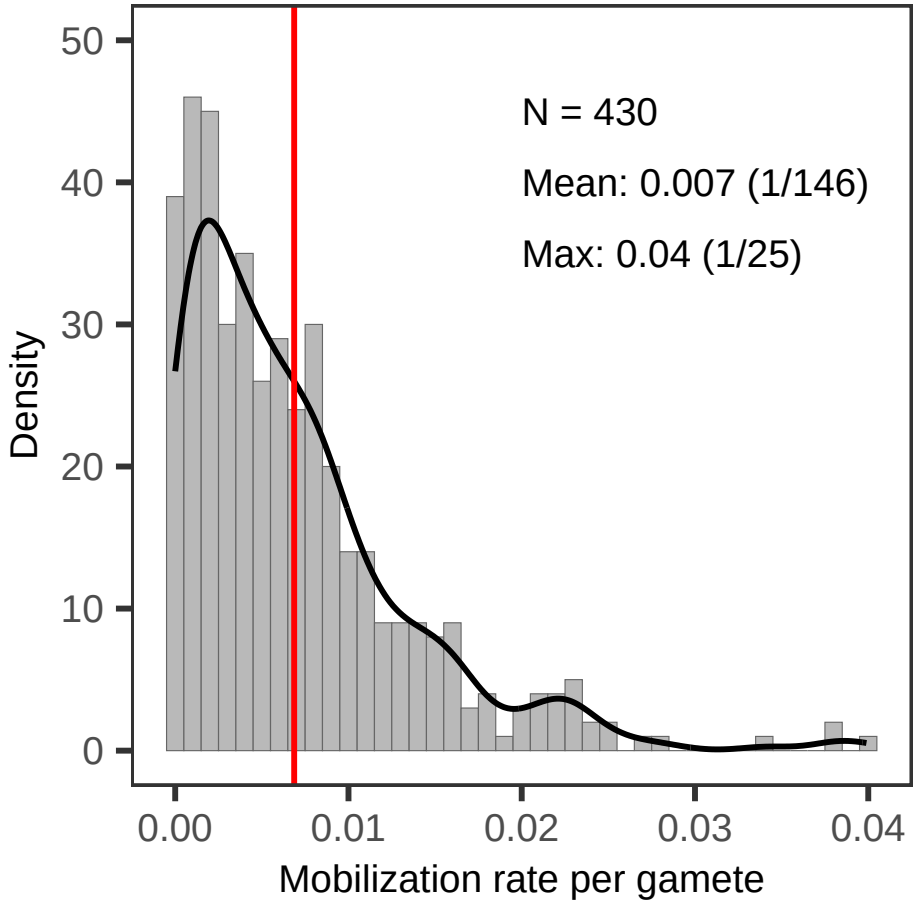
The ERVK mobilization rate in the male germline varies between individuals



3669 *de novo* ERVK

306 polymorphic constitutive ERVK





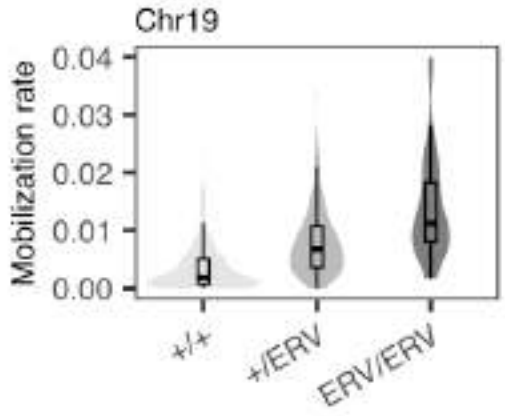
// Damona rate
~1/50
~1/125

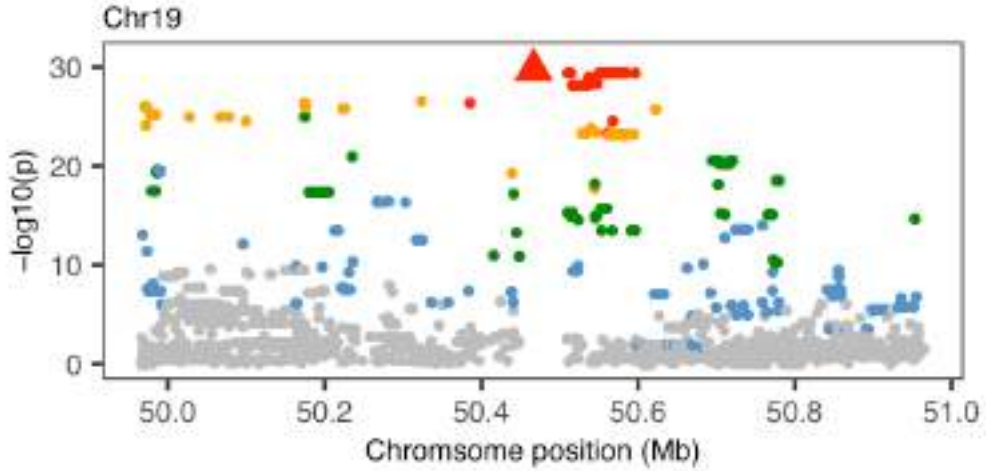
= new molecular phenotype

GWAS identifies loci affecting ERVK mobilization rate

Phenotypes:
430 *dnTR*

Genotypes:
Imputation to WGS - 10M
+ ERVK genotypes for 306 loci





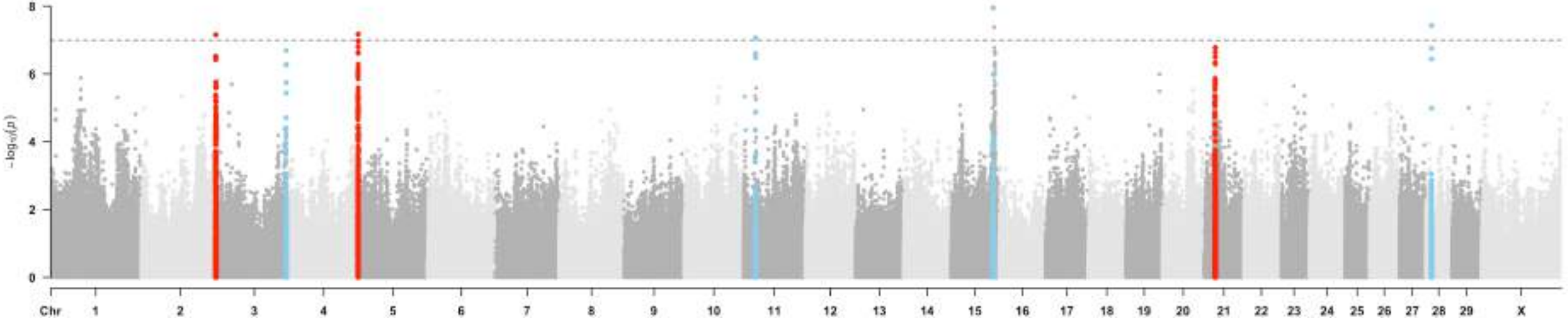
DAF = 0.35
b = 0.0052
p = $2.0e^{-30}$

GWAS identifies loci affecting ERVK mobilization rate

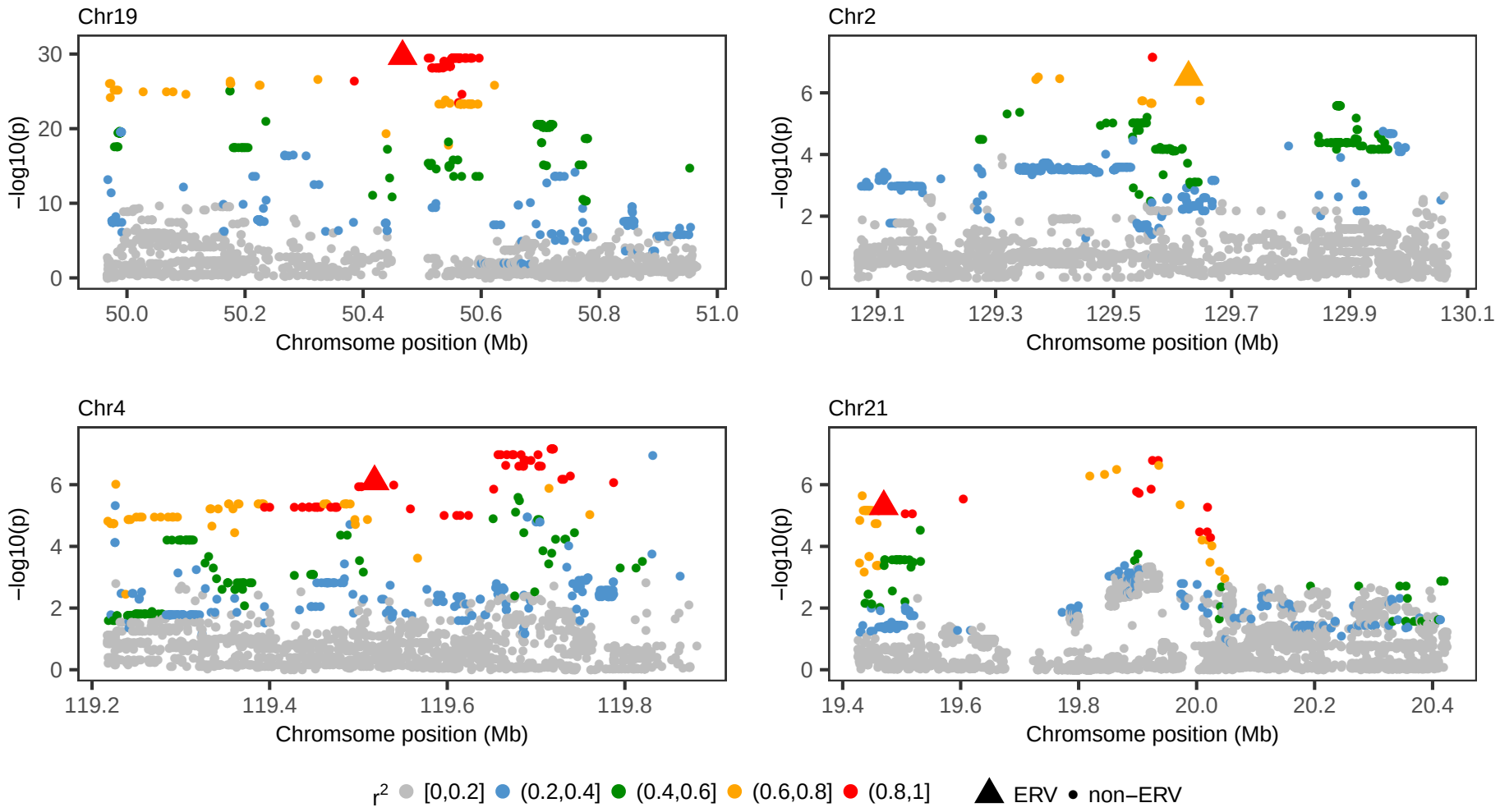
Chr19 effect fitted as covariate in the model

-> new GWAS

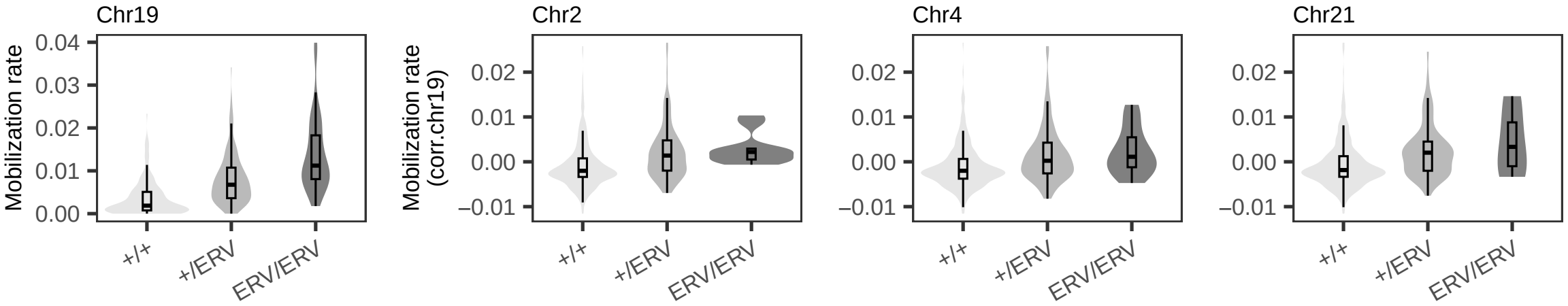
-> seven new peaks



Polymorphic ERVK elements explain four out of 8 GWAS signals



Polymorphic ERVK elements explain four out of 8 GWAS signals



DAF = 0.35
b = 0.0052
p = $2.0e^{-30}$

DAF = 0.15
b = 0.0028
p = $3.1e^{-07}$

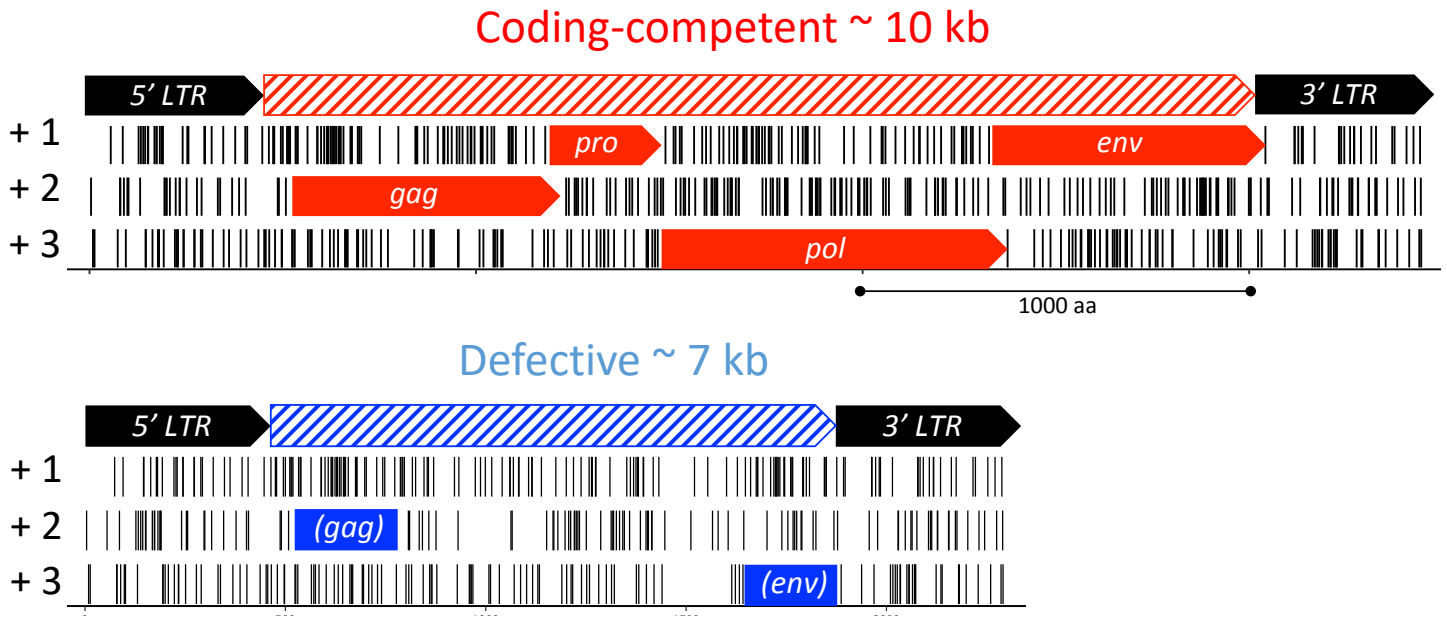
DAF = 0.22
b = 0.0023
p = $7.9e^{-07}$

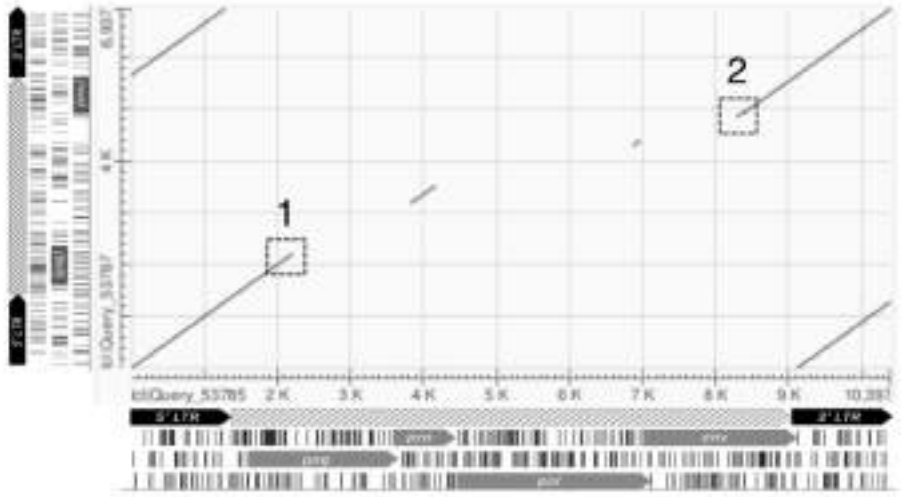
DAF = 0.11
b = 0.0029
p = $5.1e^{-06}$

What differentiates the ERVK elements in the four associated loci from the other (306 - 4) ERVK polymorphic loci?

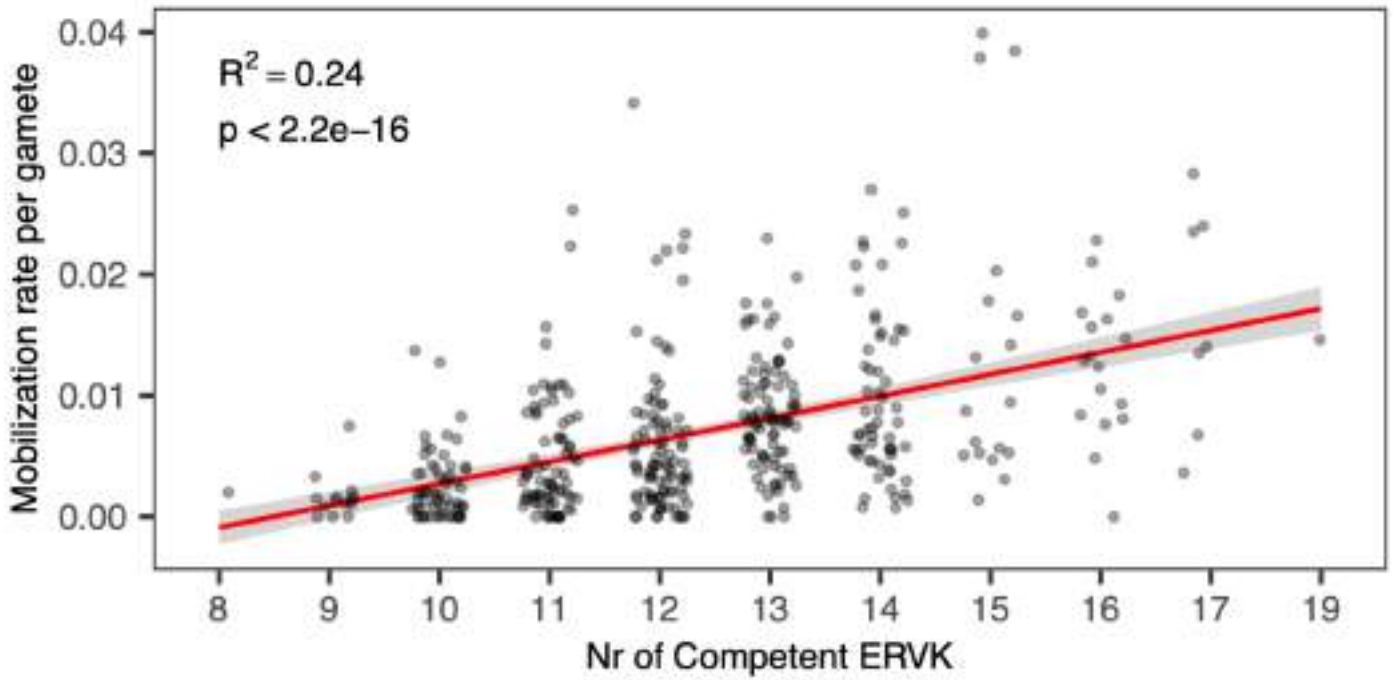
The ERVK polymorphic catalog encompasses coding-*C*ompetent and *D*efective elements

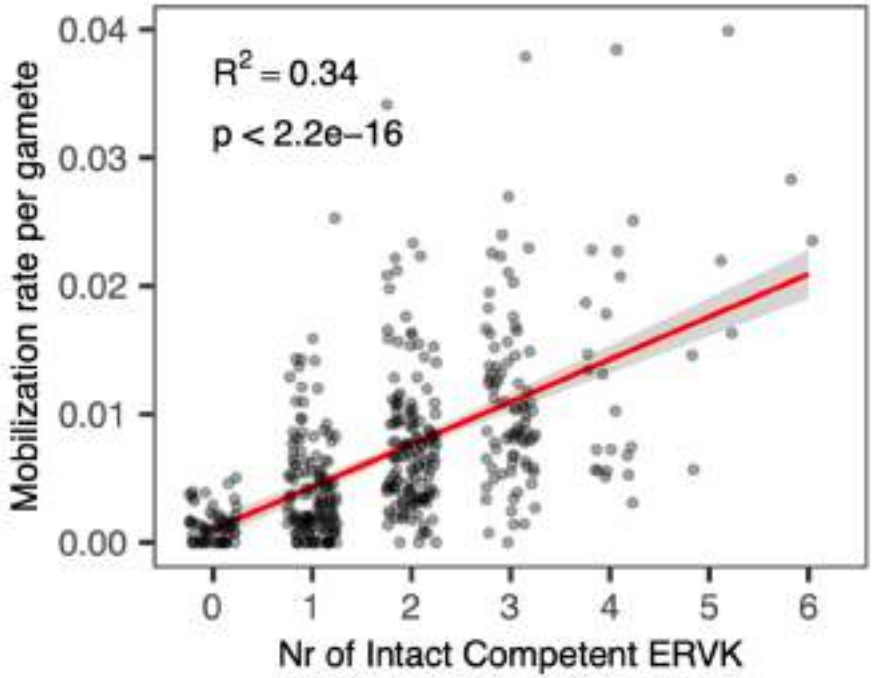
- 430 bulls, 306 polymorphic ERVK loci captured by PCIP
- PCR amplified and MiSeq sequenced: 229 (full) + 77 (half) ERVK elements
- 2 major clades: Competent (n = 50, 15%) – *Defective* (n = 256, 85%)



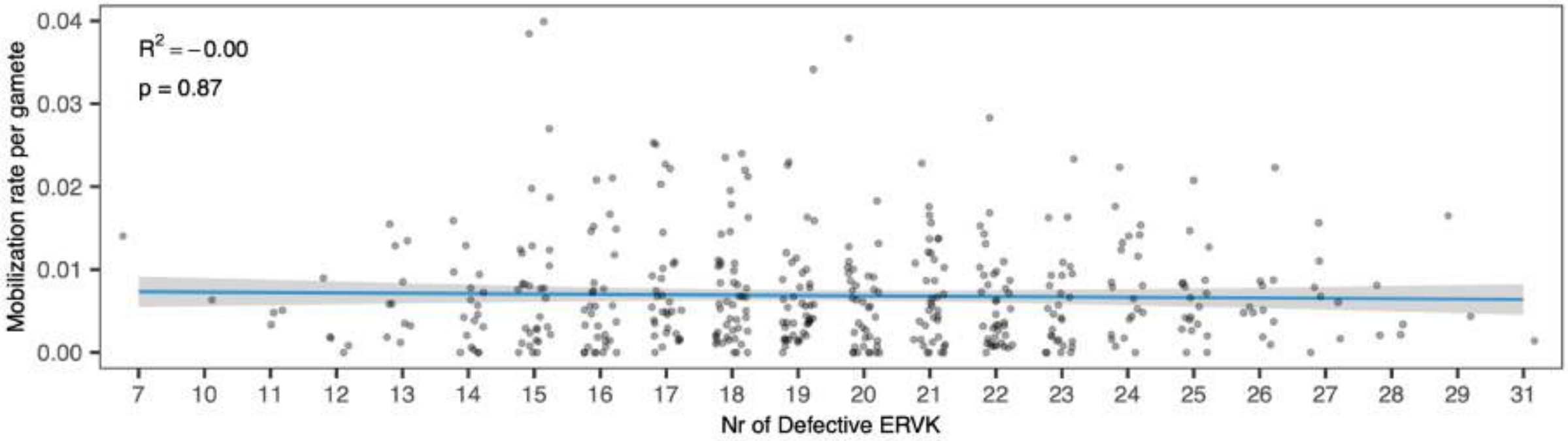


Allelic load in coding-*C*ompetent ERVK alleles explain a quarter of the *dnTR* variation





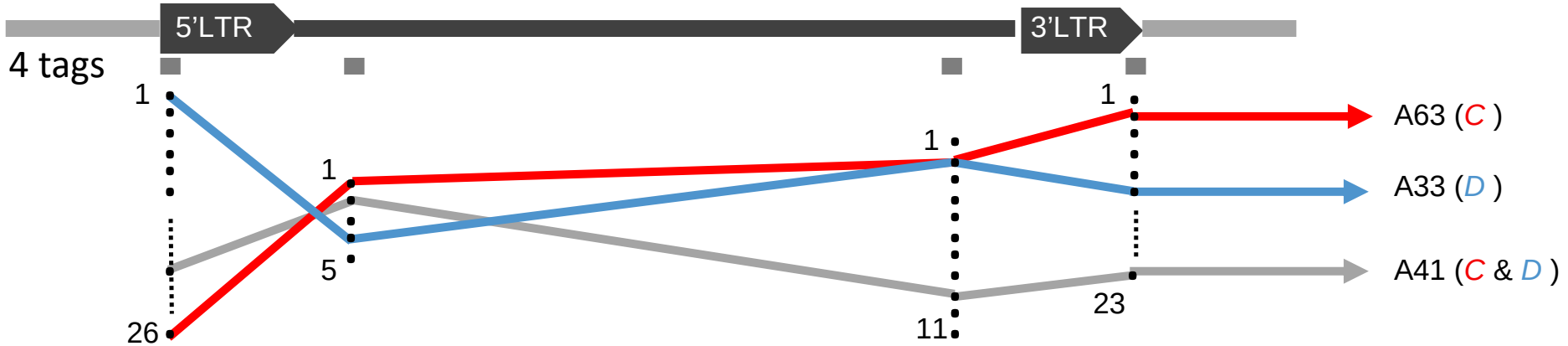
The number of non-competent clade *D* elements had no effect at all on ERVK mobilization rate



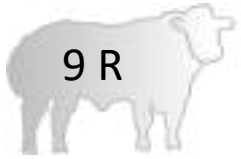
De novo ERVK insertions originate from both coding-*C*ompetent and *D*efective elements

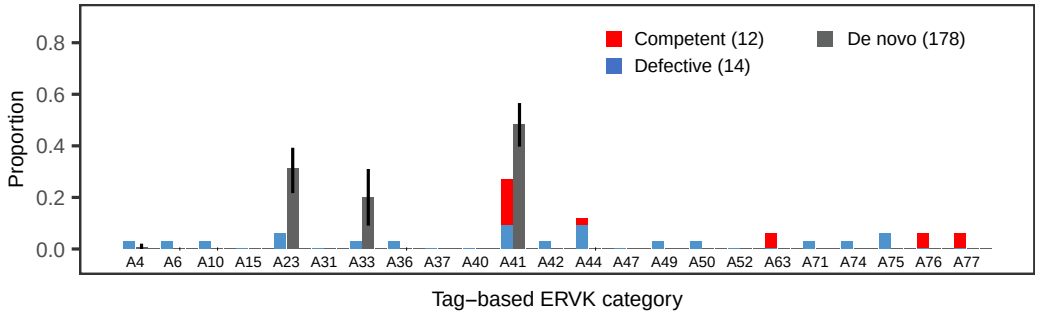
- The disease-causing insertion in *APOB* is *D*efective
- The 5 ‘pedigree-based’ *de novo* ERVK insertions are *D*efective

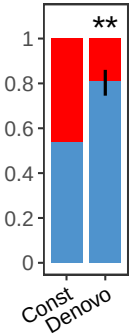
- Can we categorize *de novo* insertions as *C*- or *D*-element using PCIP ?
- Can we directly identify the donor loci ?

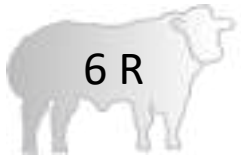


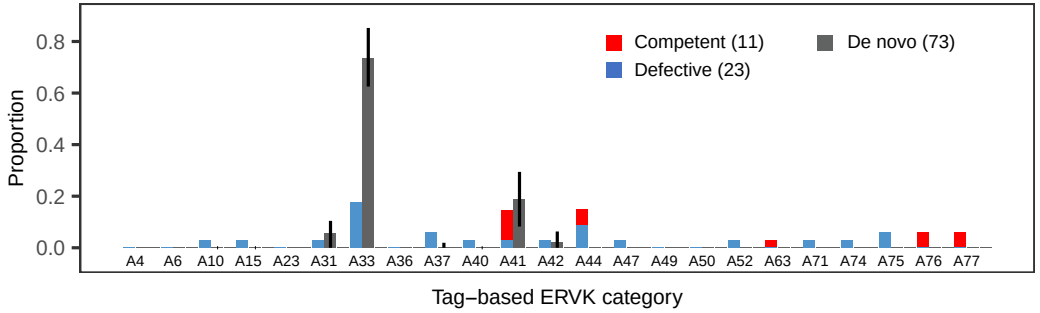
A subset of *D*efective ERVK elements are preferential donors for *de novo* insertions

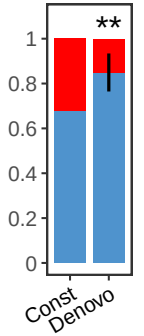


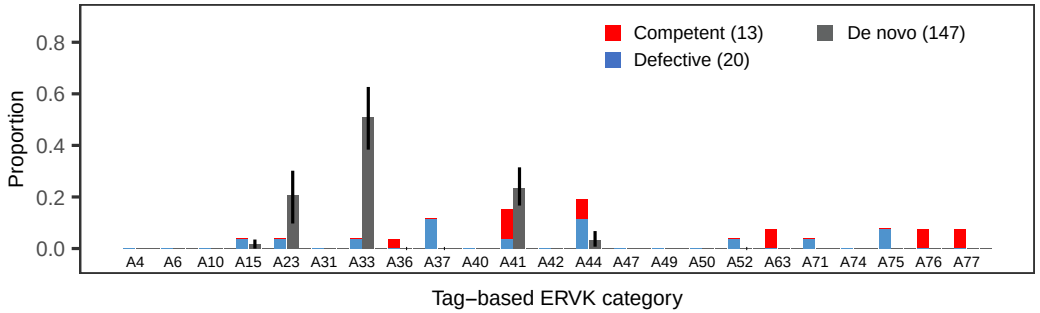


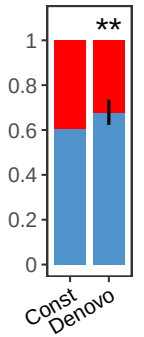




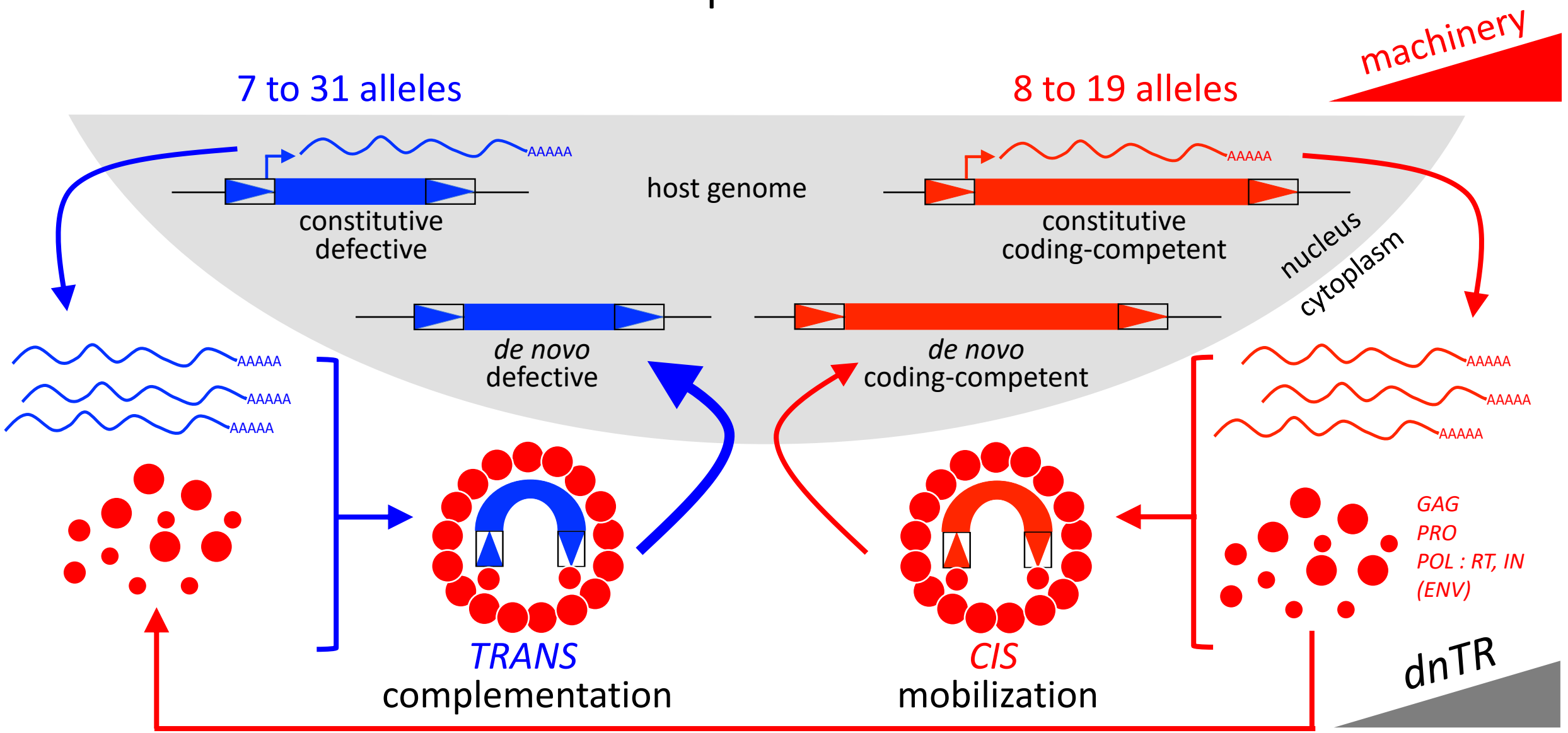








*D*efective *de novo* ERVK insertions support a trans-complementation model



Take home messages

- ERVs are still mobile in the cattle germline (both in male and female)
- *C*-elements directly influence their own *de novo* mobilization rate
- ERV mobilization rate is an actionable phenotype: for/against?
- *D*-elements are taking over by hijacking the machinery of *C*-elements
- There is no evidence from GWAS for emerging silencing mechanisms
- Maybe because ERVs are slowly self-silencing themselves (‘suicide’)

BovReg *PARTNERS*

Thank you for your attention

www.bovregproject.eu

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668