

The life history of retrocopies illuminates the evolution of new mammalian genes

Francesco Nicola Carelli, Takashi Hayakawa, Yasuhiro Go, Hiroo Imai,
Maria Warnefors, and Henrik Kaessmann.

Genome Research, Advance ePub, January 4th

Retrocopies

A

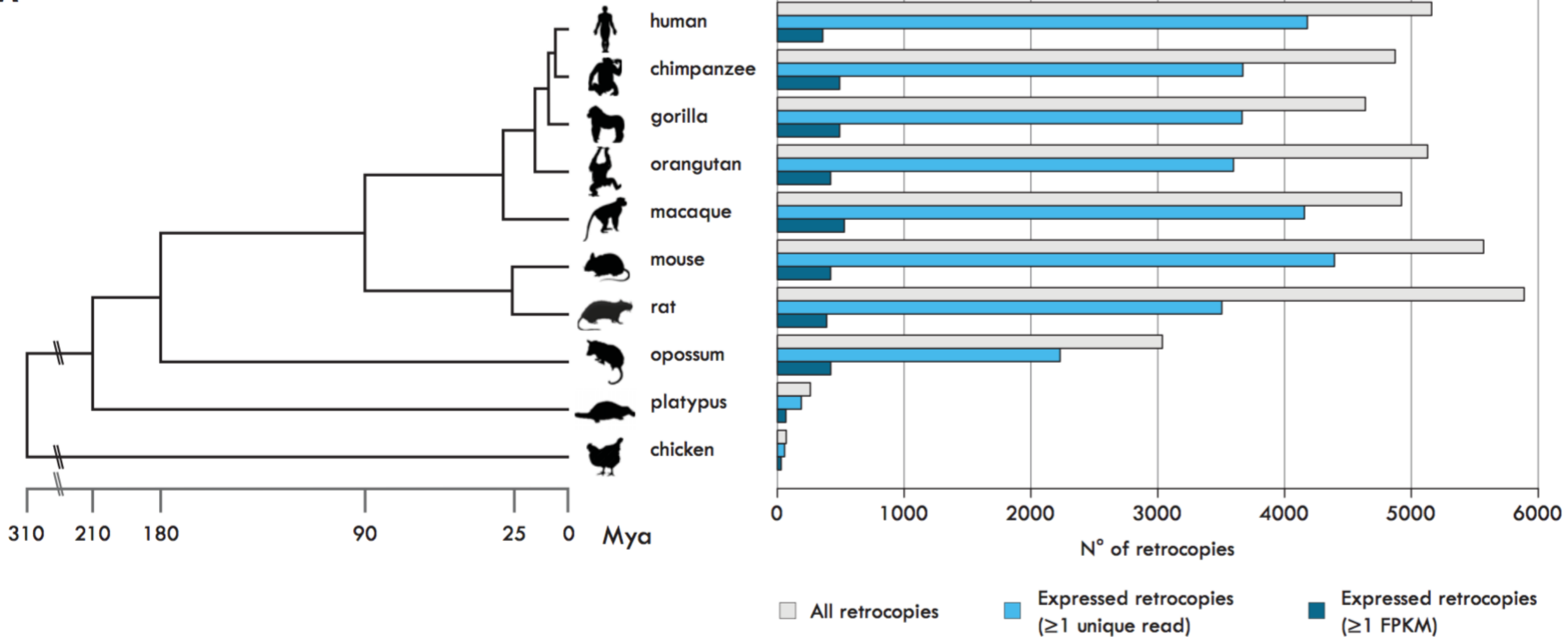
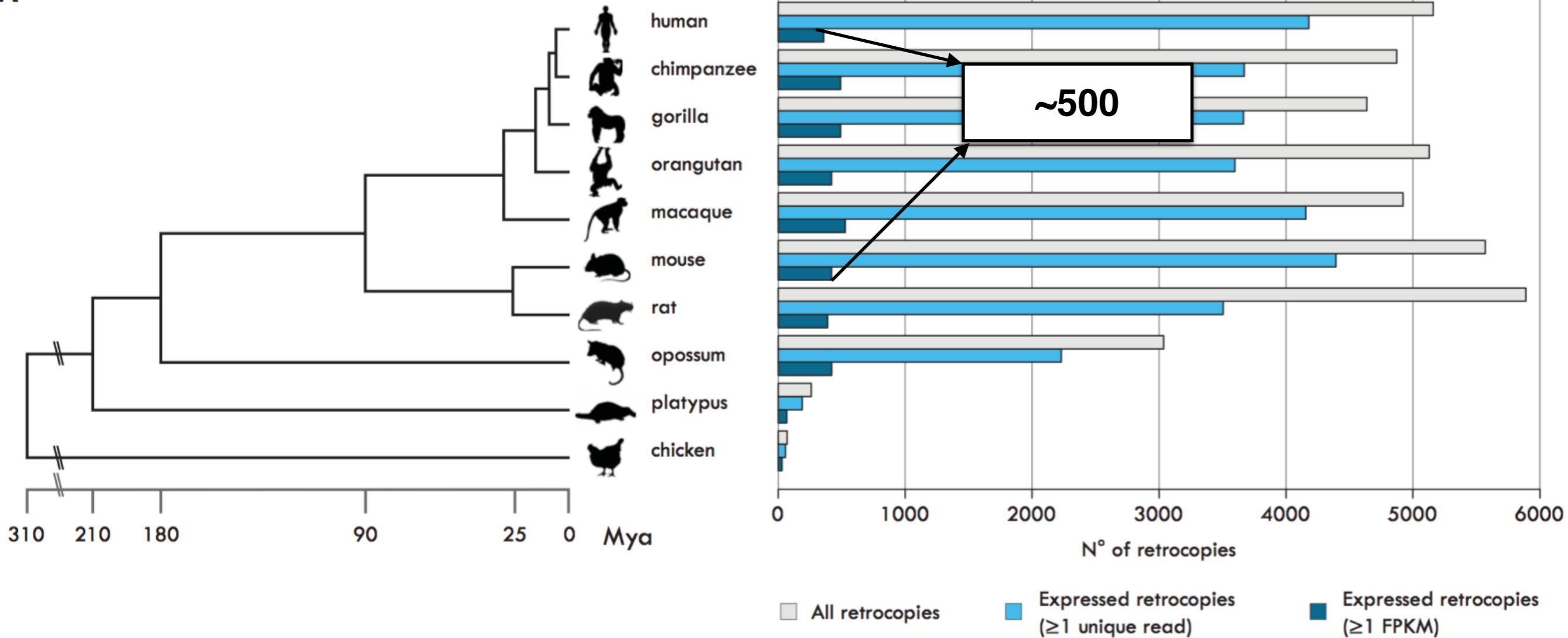


Fig 1

Retrocopies

A



GENCODE v24

Pseudopipe

Retrofinder

RCPedia

Processed pseudogenes

10,283

8,739

~15,000

7,831

3

Fig 1

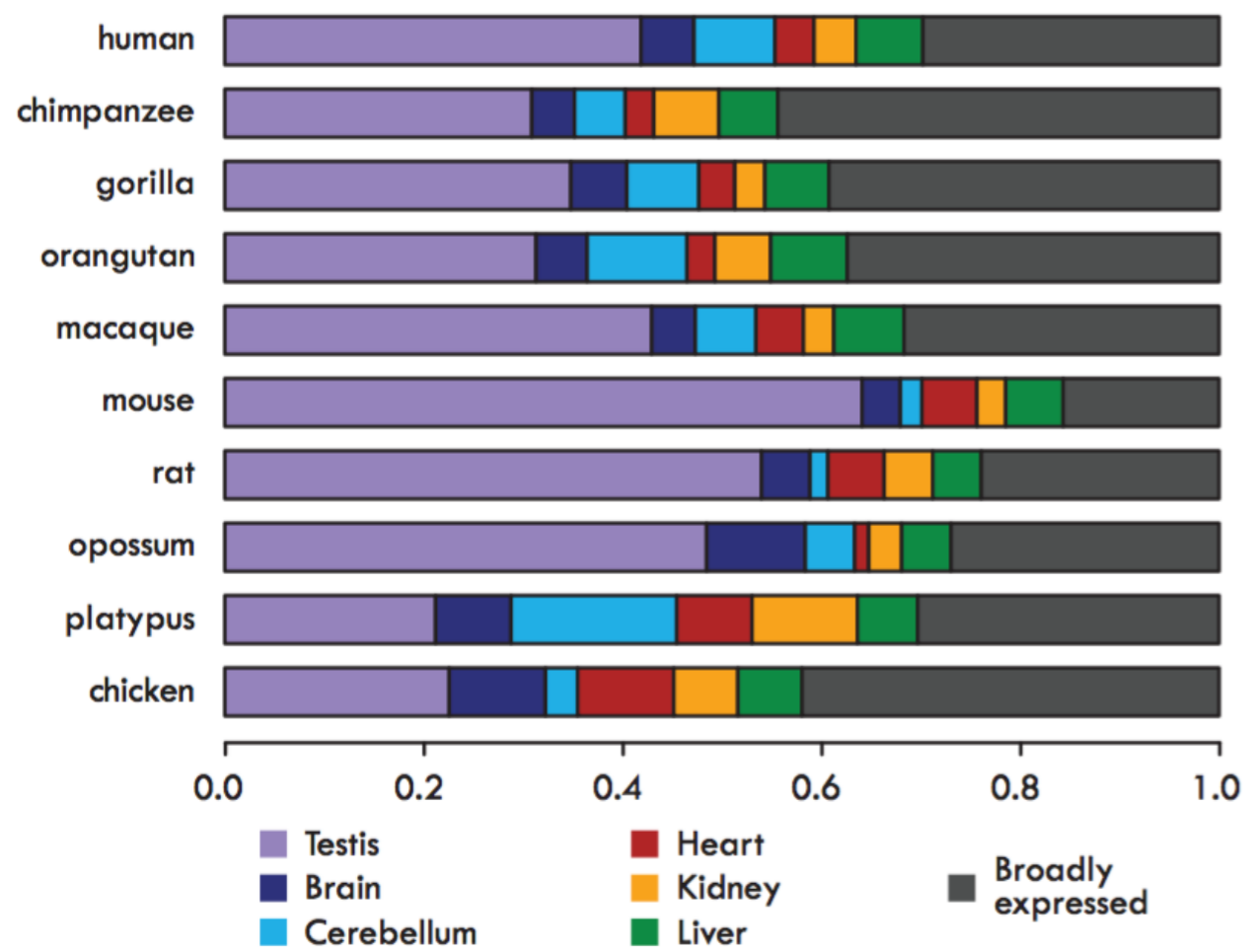
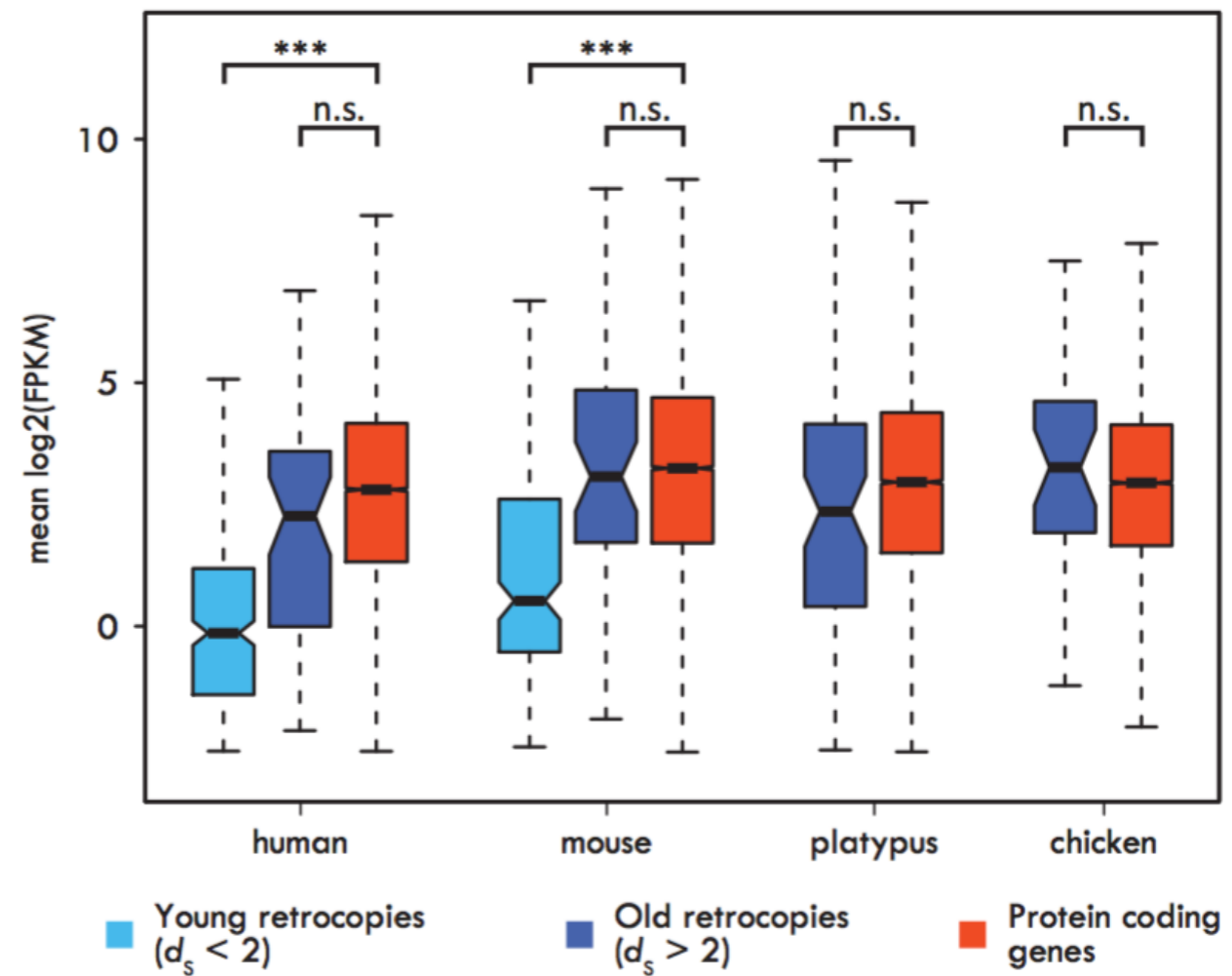
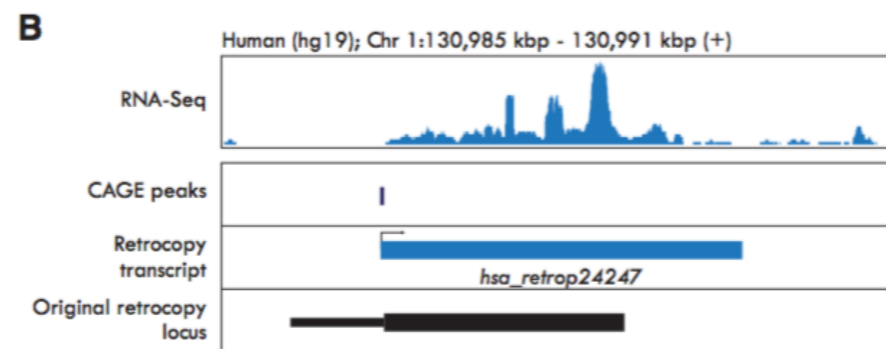
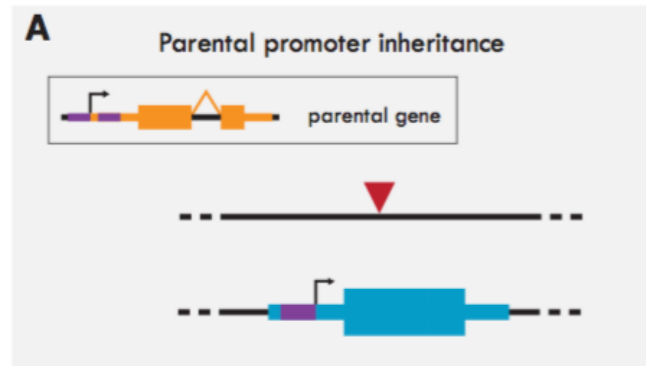
B**C**

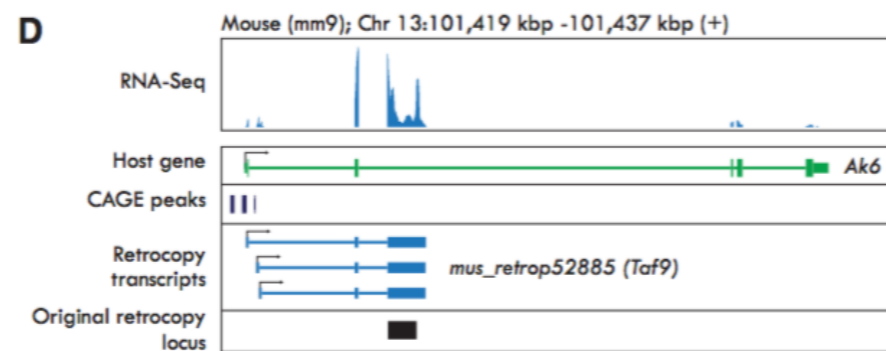
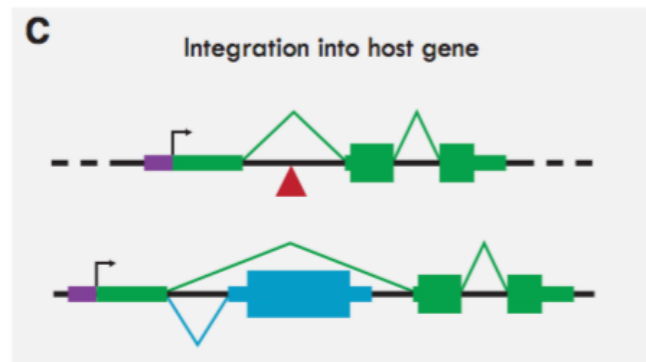
Fig 1

230 human and 243 mouse high-confidence TSS (CAGE)

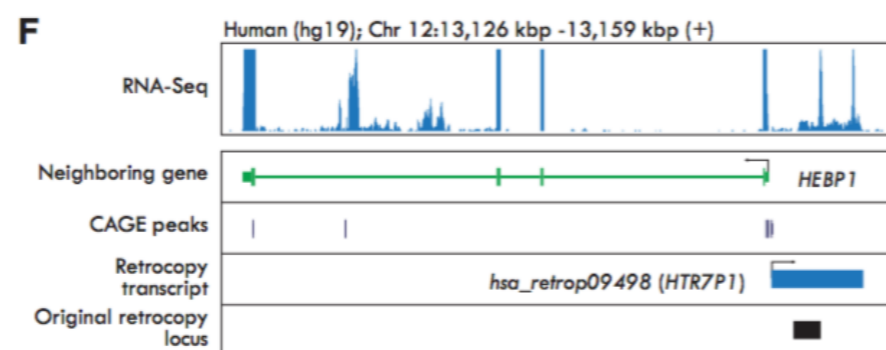
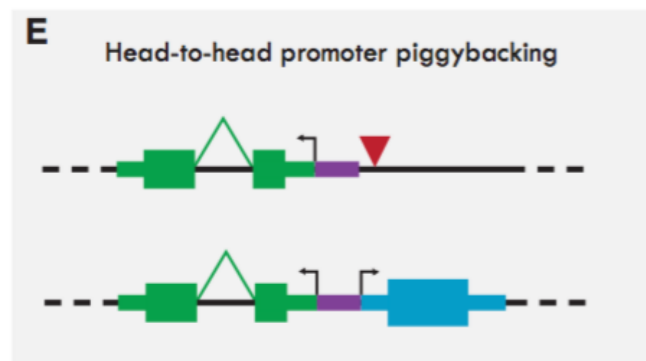
Fig2



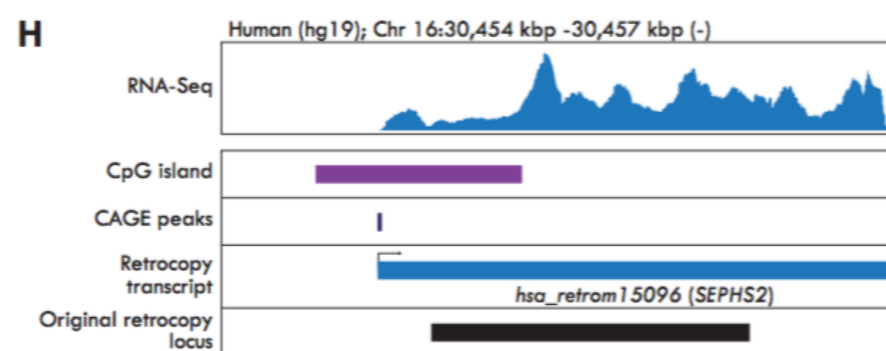
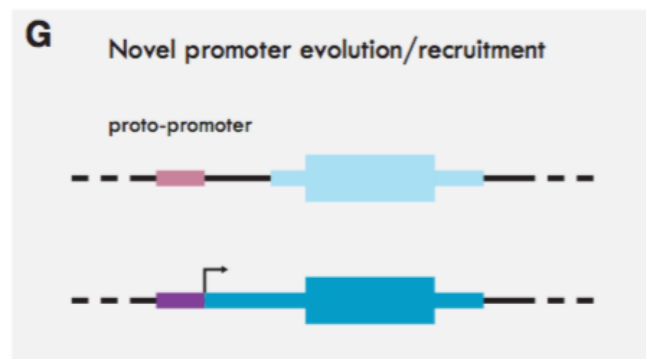
~3% inherited promoters from parental sequences



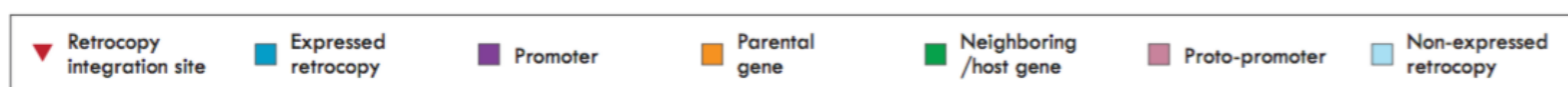
~11% acquired promoters from host genes



~5% acquired promoters from head-to-head



~50% CpG islands



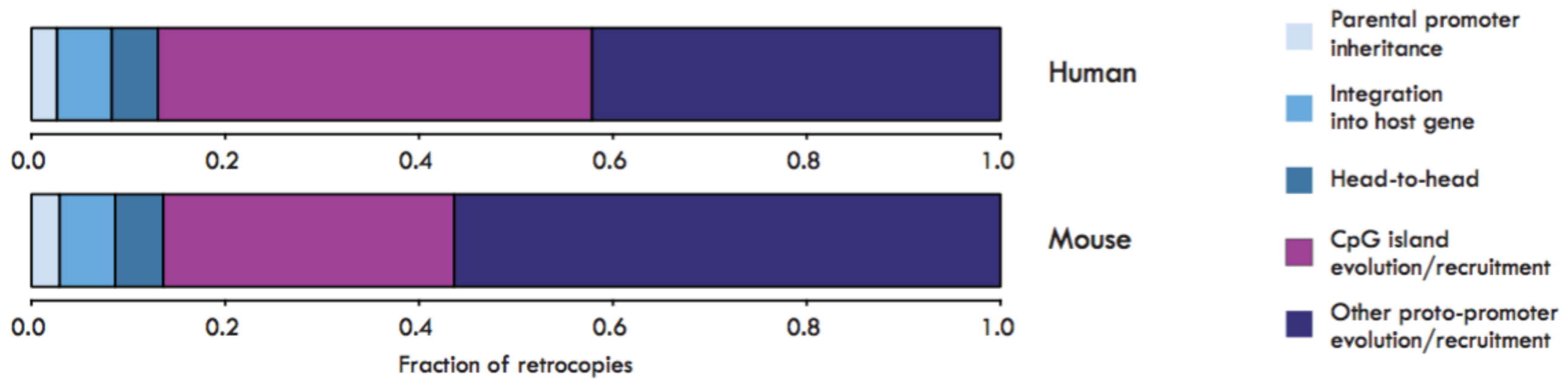


Fig2

Porto-promoters / Enhancers

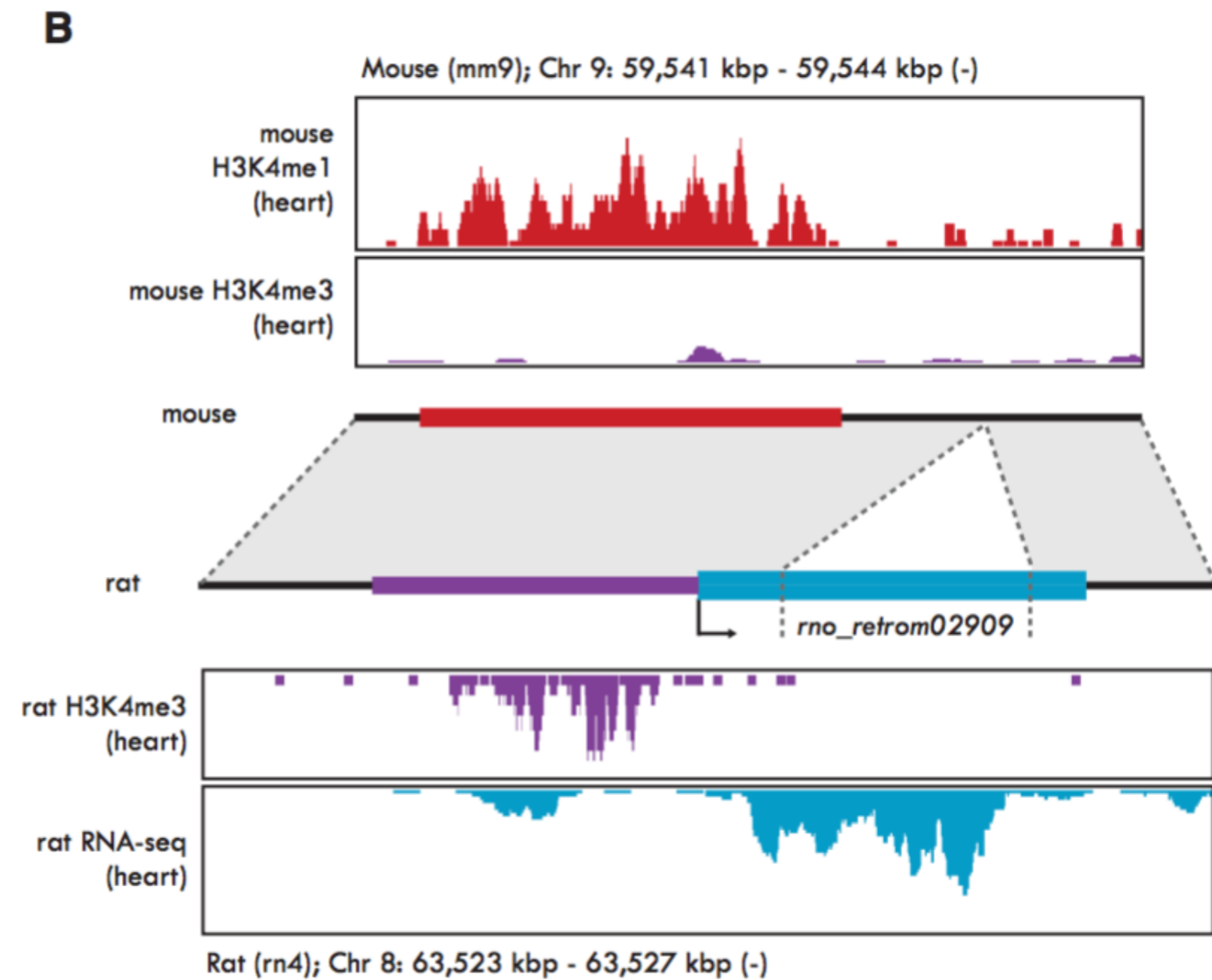
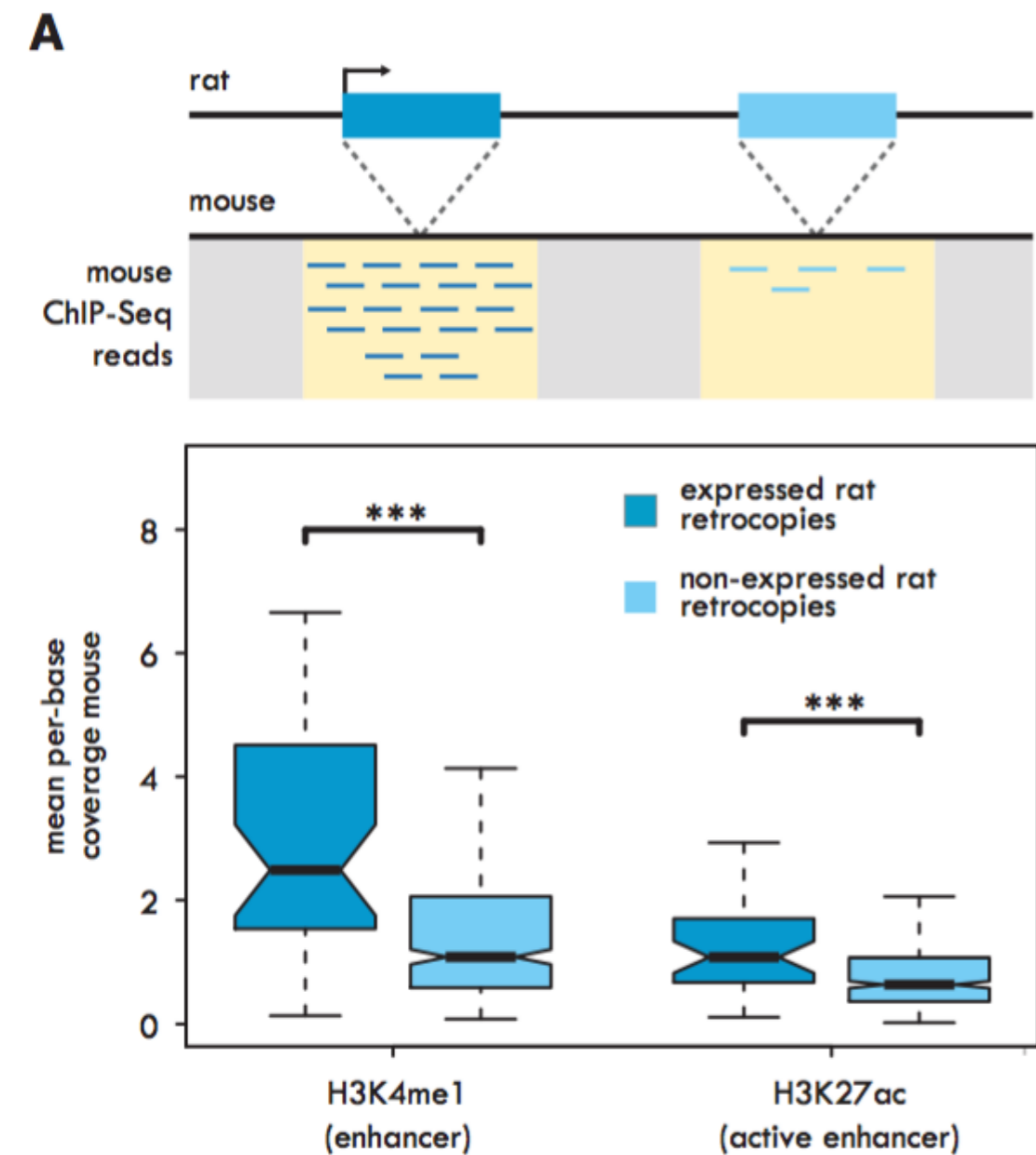


Fig3

~40% acquired promoters from proto-promoters (enhancers)

Retrogenes

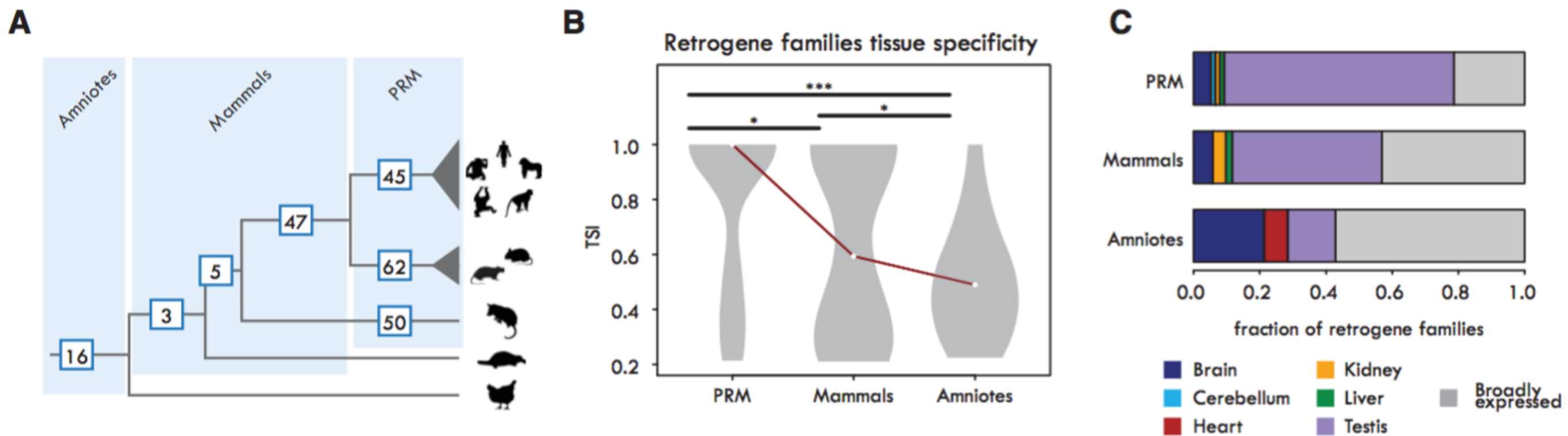


Fig4

Out of X Hypothesis

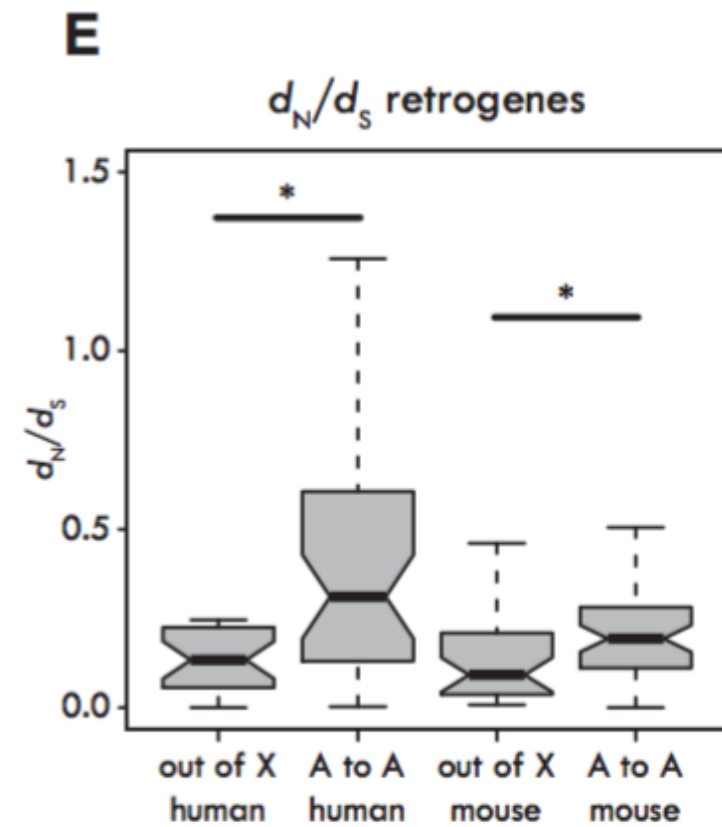
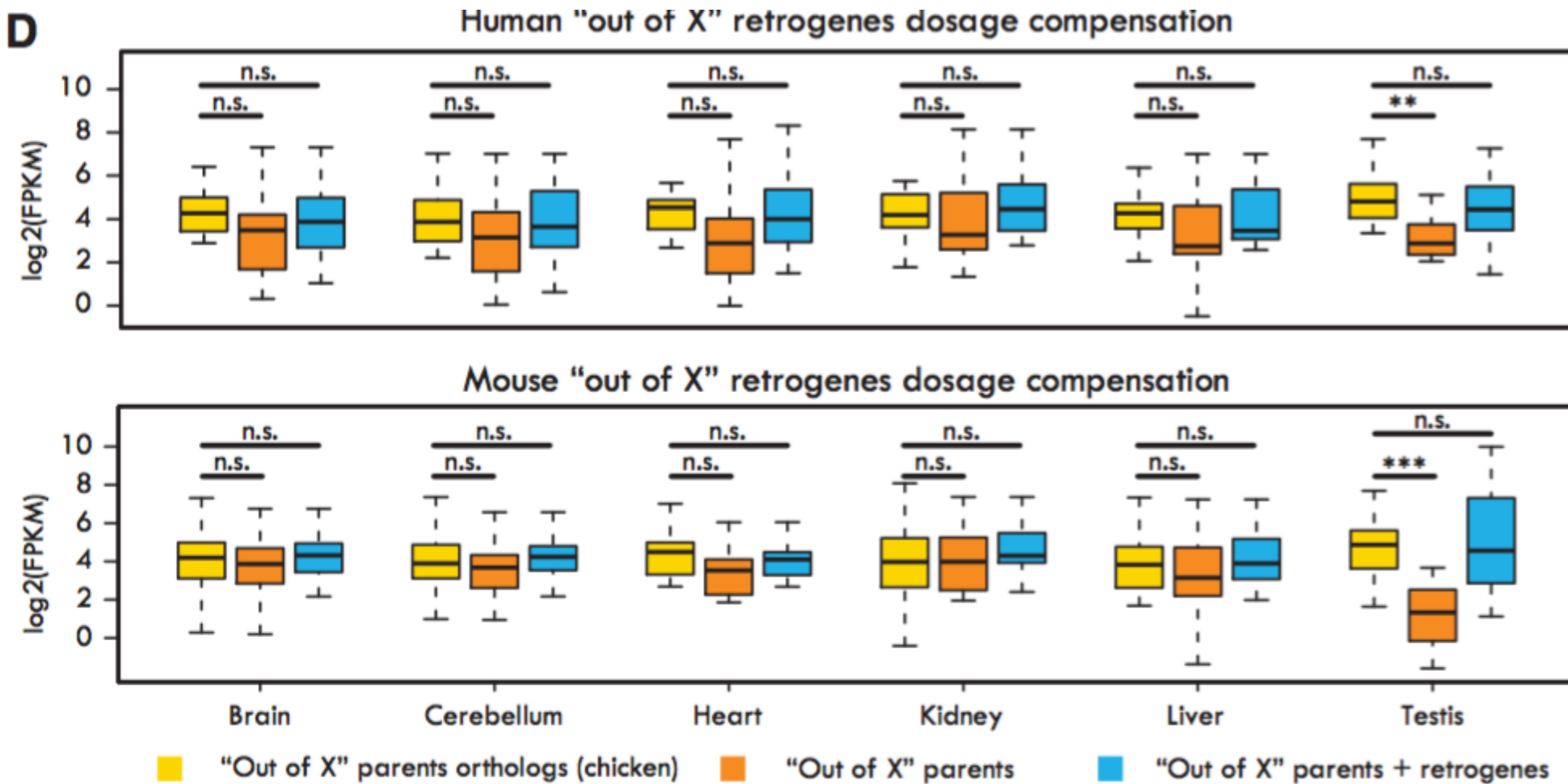


Fig4

Table 1. Features of vertebrate retrogenes

	Retrogene no.^a	Multiexonic retrogenes^b	Intronization events^c	New 5' exons^c	New 3' exons^c	Retrogenes alternative transcripts^b
Human	199 (117)	48 (41%)	3 (6%)	41 (85%)	11 (22%)	47 (40%)
Chimpanzee	224 (122)	30 (25%)	0 (0%)	25 (83%)	9 (30%)	32 (26%)
Gorilla	215 (120)	38 (32%)	2 (5%)	32 (84%)	7 (18%)	37 (30%)
Orangutan	194 (102)	33 (32%)	1 (3%)	29 (87%)	7 (21%)	65 (63%)
Macaque	198 (122)	46 (38%)	2 (4%)	39 (84%)	11 (23%)	51 (41%)
Mouse	213 (132)	49 (37%)	0 (0%)	42 (85%)	14 (28%)	64 (48%)
Rat	280 (154)	47 (30%)	0 (0%)	37 (78%)	14 (29%)	55 (35%)
Opossum	256 (171)	48 (28%)	1 (2%)	39 (81%)	13 (27%)	48 (28%)
Platypus	88 (40)	16 (40%)	1 (6%)	12 (75%)	6 (37%)	8 (20%)
Chicken	36 (28)	15 (53%)	0 (0%)	14 (93%)	6 (40%)	10 (35%)

^aNumber of expressed retrogenes is in parentheses.

^bFraction of expressed retrogenes multiexonic or with alternative transcripts is in parentheses.

^cFraction of multiexonic retrogenes is in parentheses.

Alternative splicing

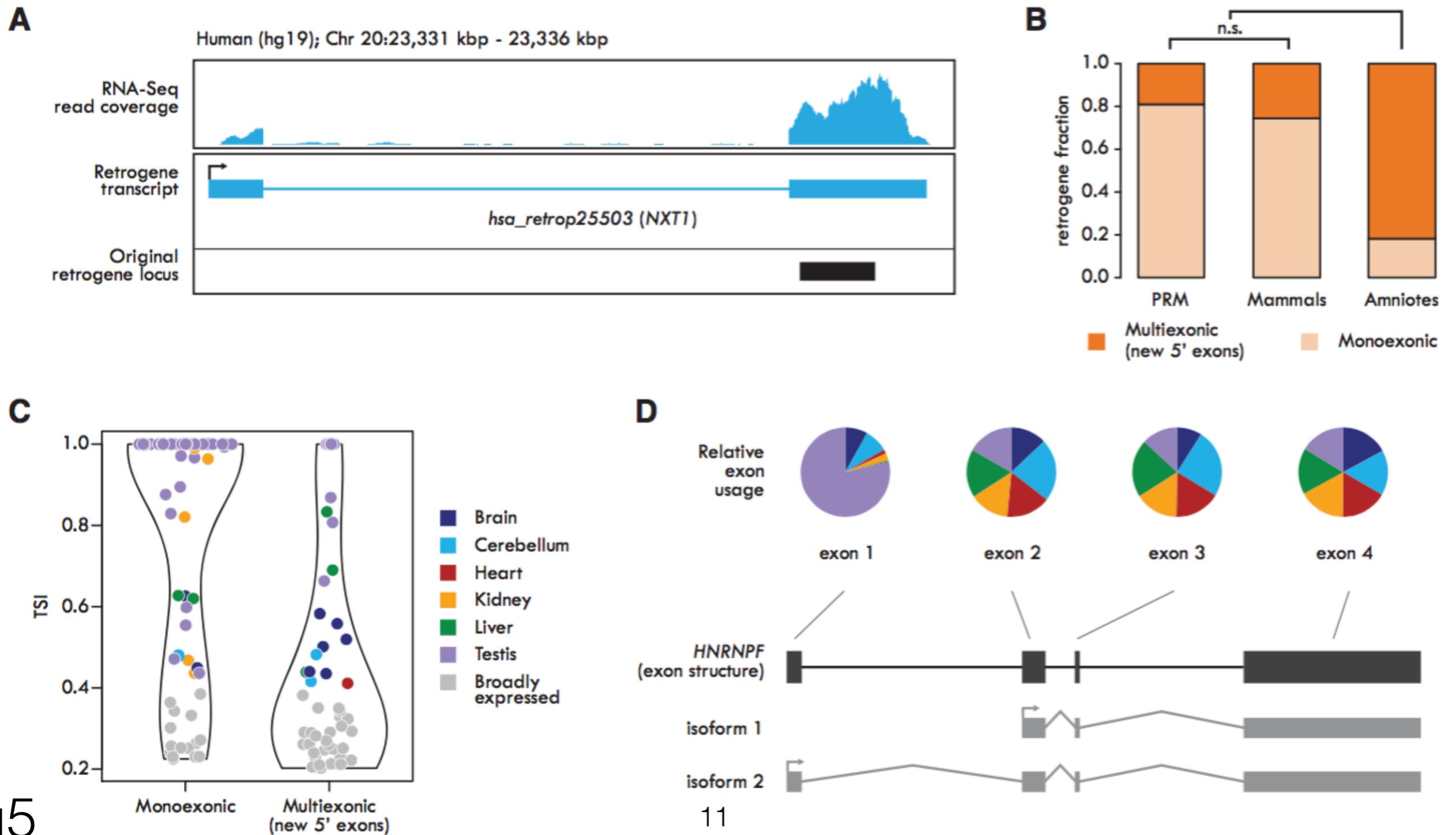


Fig5

Table 2. Orphan retrogenes

Retrogene name	Clade	ED ^a	Function	Parental gene replacement	Promoter source ^b	Structure
<i>DEM1</i>	Eut	8.9	Single-stranded DNA exonuclease	—	CGI; HH	Multiexonic
<i>CNO</i>	Eut	23.98	Part of BLOC-1 complex (organelle biogenesis)	—	CGI	Monoexonic
<i>TRMT12</i>	Eut	6	Enzyme involved in wybutosine synthesis	Human gene rescues yeast parental KO (Rodriguez et al. 2012)	Novel	Monoexonic
<i>HYI</i>	Mar	52.15	Hydroxypyruvate isomerase	—	CGI	Monoexonic
<i>L2HGDH</i>	Mar	6.82	L-2-hydroxyglutarate dehydrogenase	—	CGI	Monoexonic
<i>ZNF830</i>	Mam	5.68	Zinc-finger protein (embryo development)	—	CGI; HH	Monoexonic
<i>LCMT2</i>	Mam	9.61	Enzyme involved in wybutosine synthesis	—	CGI; HH	Multiexonic
<i>COMMD5</i>	Mam	4.71	Cell proliferation control	—	CGI	Multiexonic
<i>MARS2</i>	Mam	8.07	Methionyl-tRNA synthetase 2	Human gene rescues fruitfly parental KO (Bayat et al. 2012)	CGI	Monoexonic
<i>RNF113</i>	Mam	26.11	Zinc-finger protein (RNA splicing)	Human gene rescues fruitfly parental KO (Carney et al. 2013)	CGI; HH	Monoexonic

(Eut) Eutherian-specific; (Mar) marsupial-specific; (Mam) mammalian-specific; (CGI) CpG island; (HH) head-to-head promoter recruitment.

^aEuclidean distance measured between the log₂ transformed FPKM values of orthologous orphan retrogenes and the parental genes in outgroup species.

^bThe promoter sources are referred to human or opossum (for marsupial-specific) genes; opossum CpG island coordinates were obtained from the UCSC Genome Browser website.

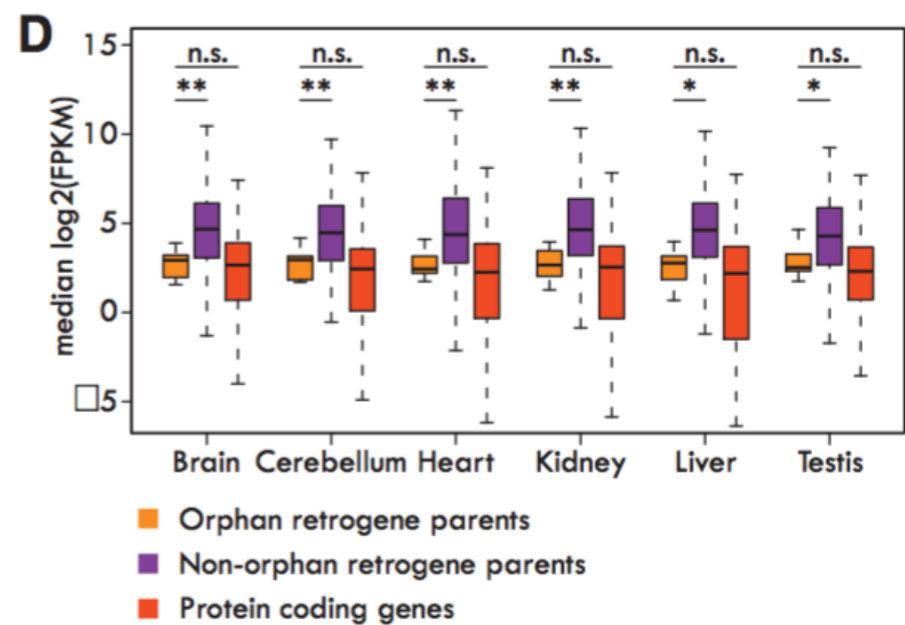
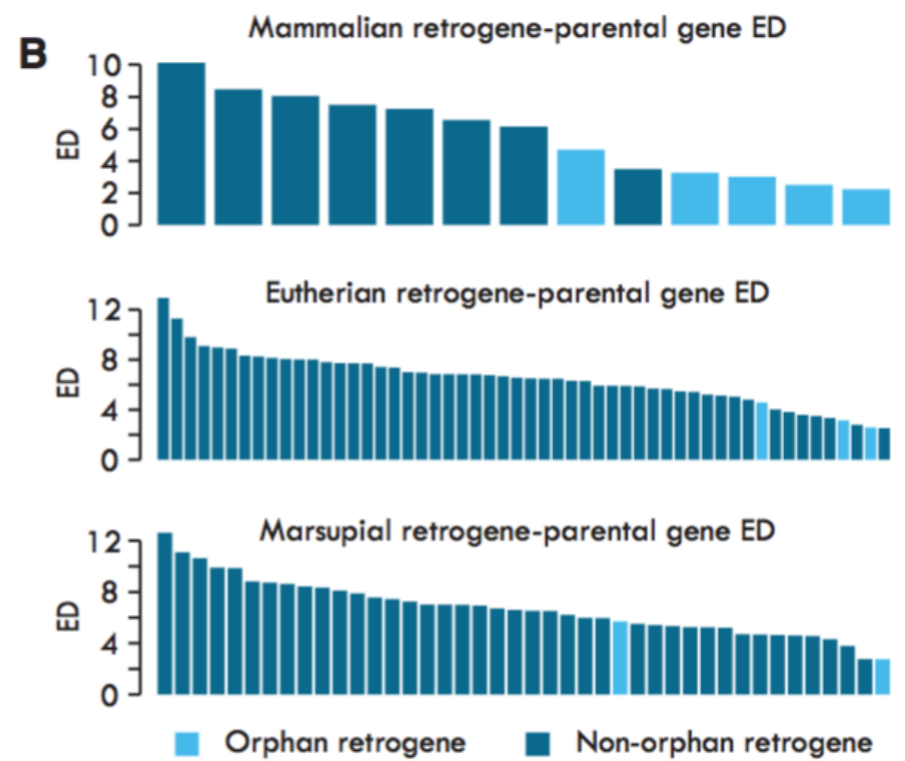
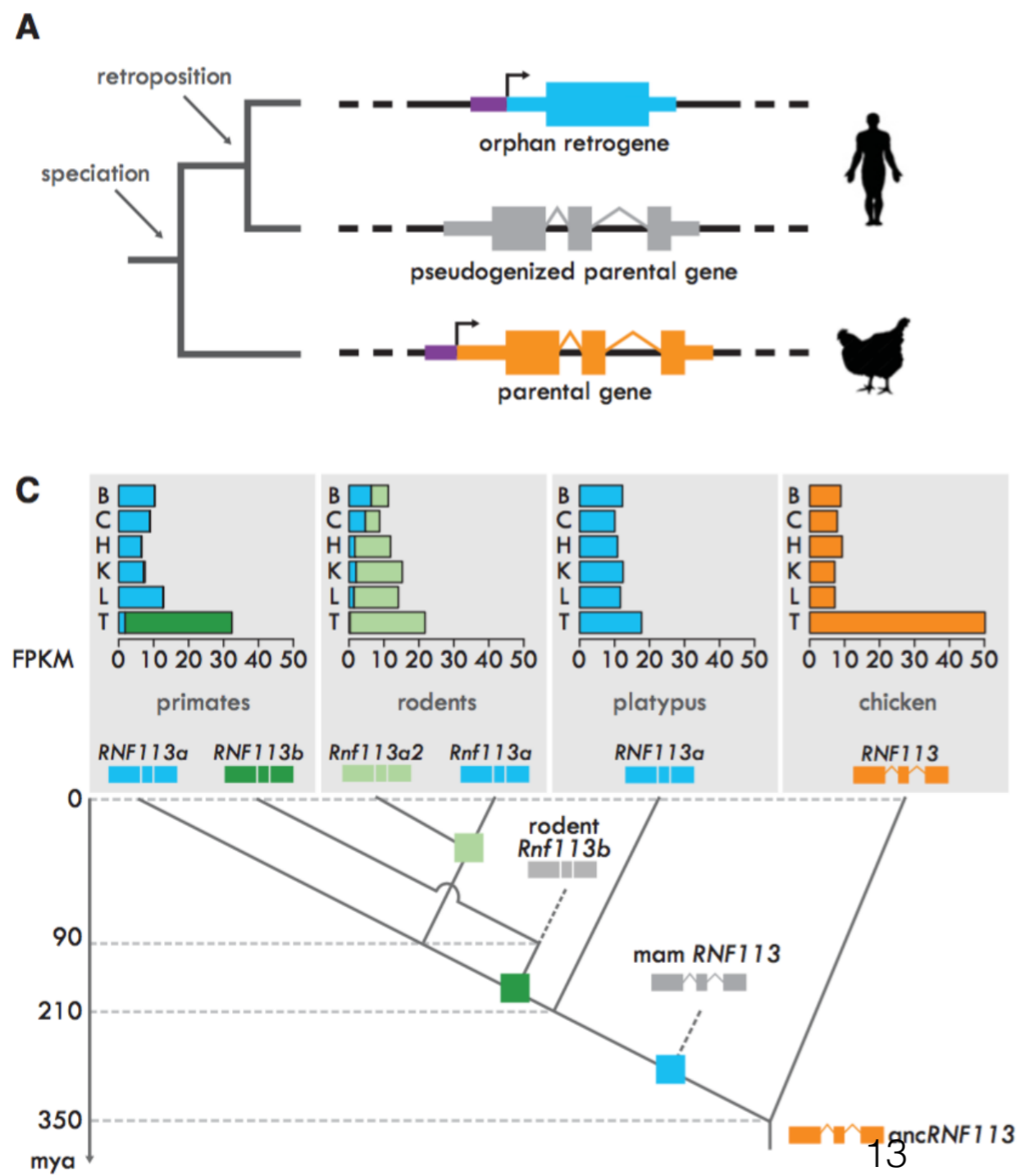


Fig5

Questions?