Bioinformatic analysis of Ruminant Endogenous Retrovirus expression

Stage volontaire:

IVPC (INRAE)-LBBE(UCBL)

Superviseurs:

Jocelyn TURPIN

Emmanuelle LERAT

Marie VERNERET

Transposable element overview

- TE are mobile genetic element.

- Represent a large part of mammalian genome.

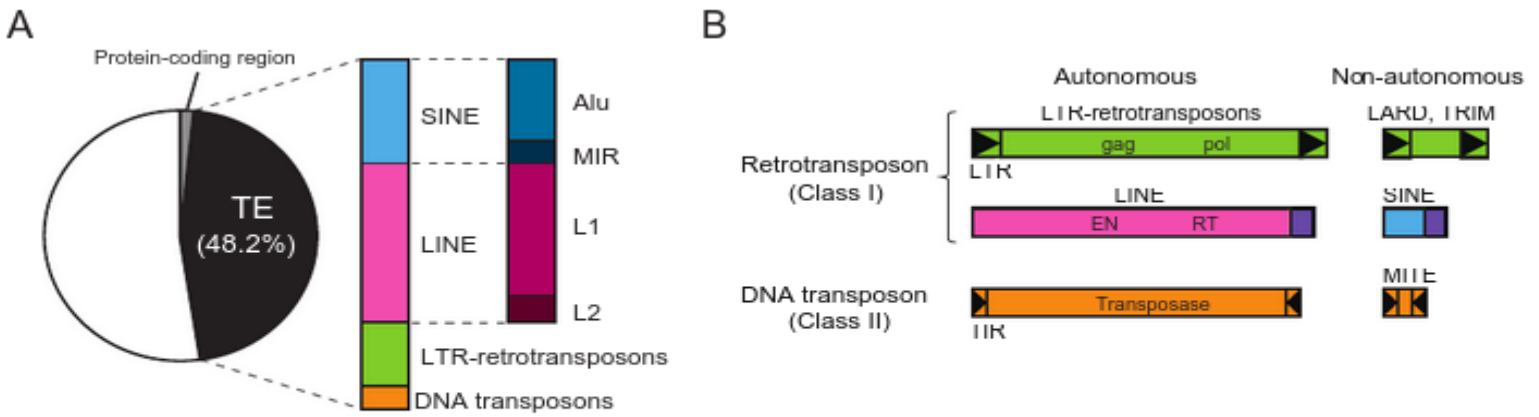


Fig. 1. Fraction of TEs in the human genome and classification of TEs. (A) Proportion of protein-coding sequences (gray), TEs (black), and other DNA (white) in the human genome. Major components of SINEs are *Alu* and MIR, whereas those of LINEs are L1 and L2. (B) Basic classifications of eukaryotic TEs. LTRs and TIRs are represented by boxes with triangles. The 3' end sequences shared between LINEs and SINEs are represented by purple boxes. EN, endonuclease; RT, reverse transcriptase.

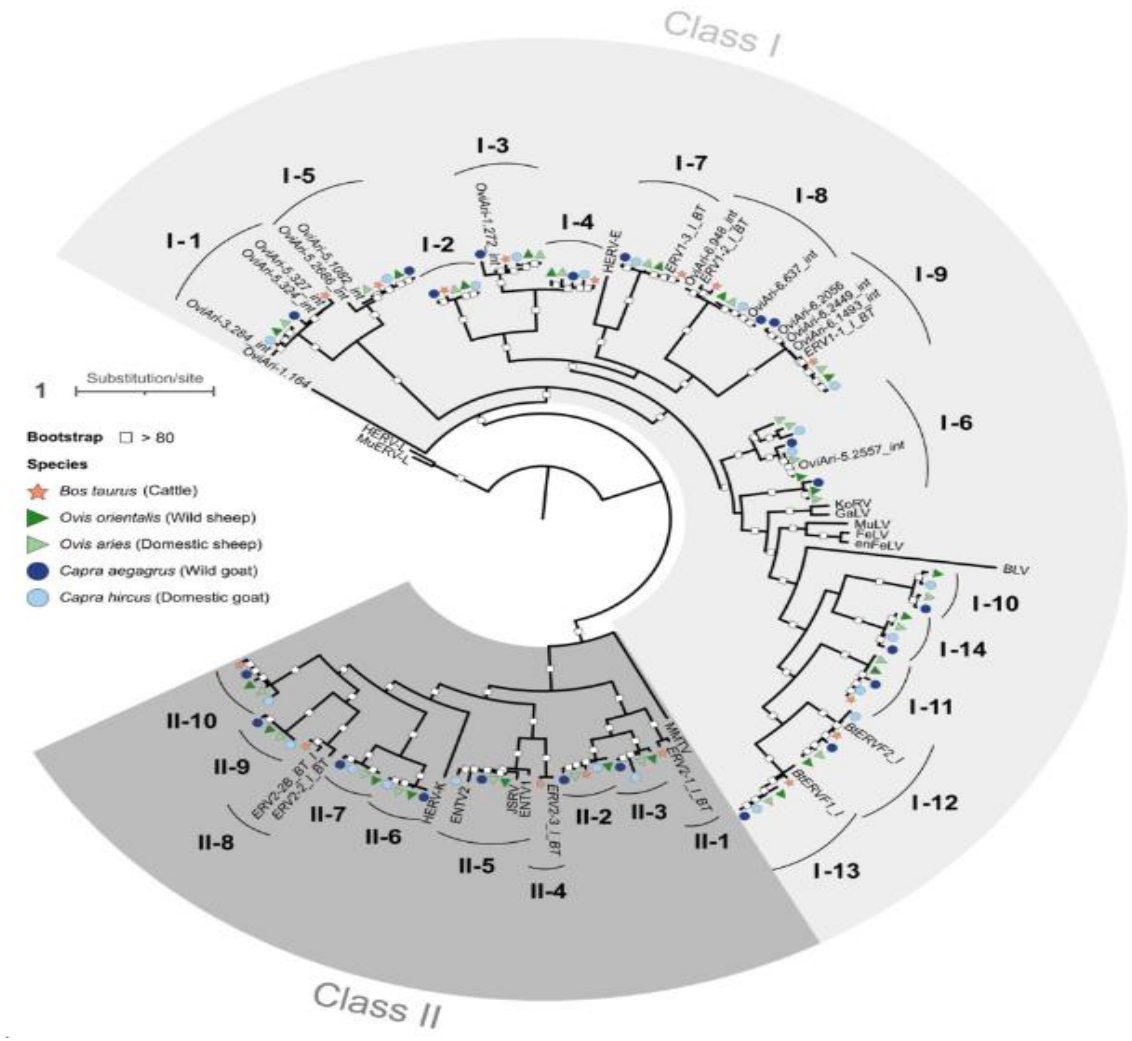
Hidenori Nishihara 2020

Retroviruses in small ruminants (plus gros caractères)

- Exogenous retroviruses (JSRV and ENTV) pathogenic.

- Endogenous retroviruses (ERV1 - class1 and ERVK - class2) have been less described; their repertoire and evolutionary history remain unknown.
- ERVII-5 related to …

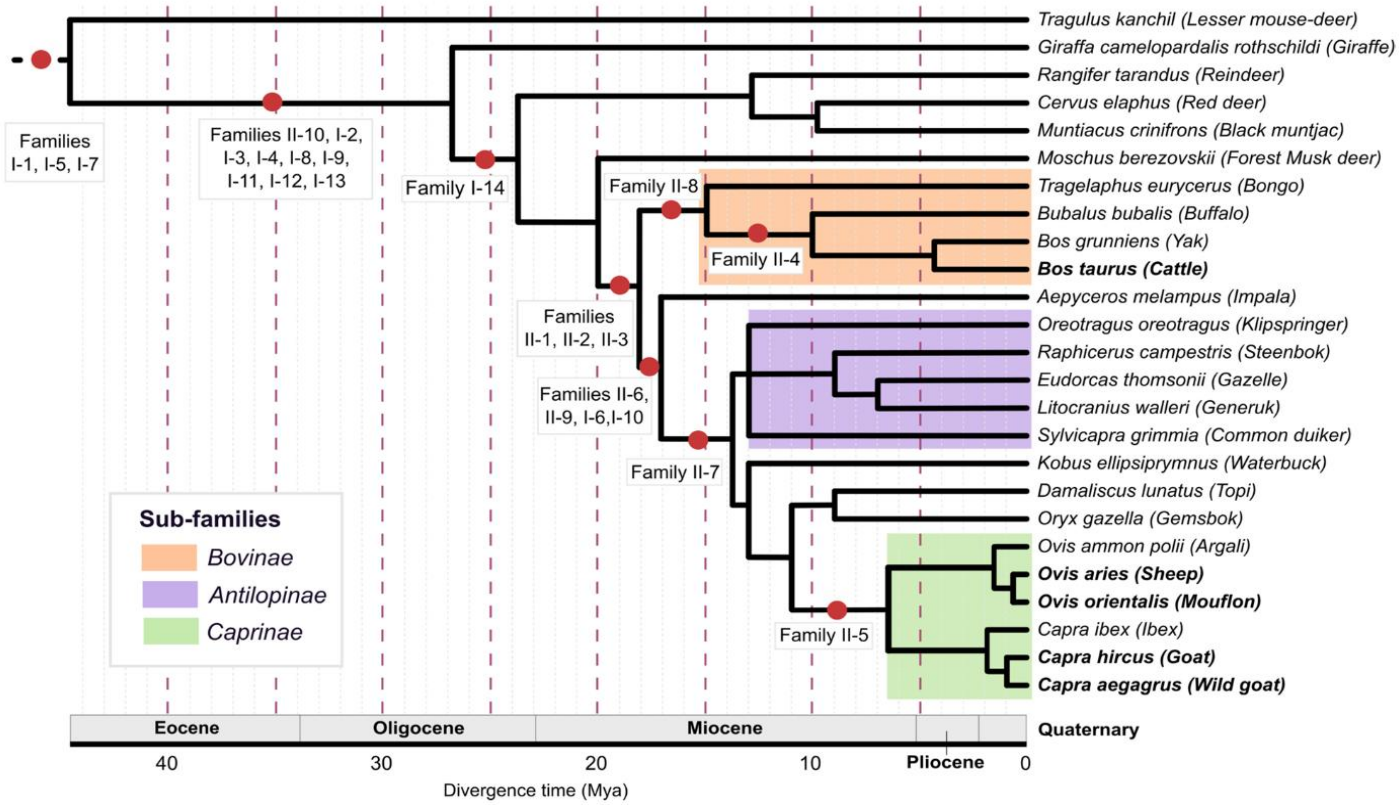
Characterization of ERVs families in ruminants



- 24 ERVs families characterized including : 14 ERV1 families and 10 ERVK families.
- ERV1 = class1.
- ERVK = class2.
- Preciser ceux qui sont spécifiques aux caprinae (21)

Verneret et al mobile DNA

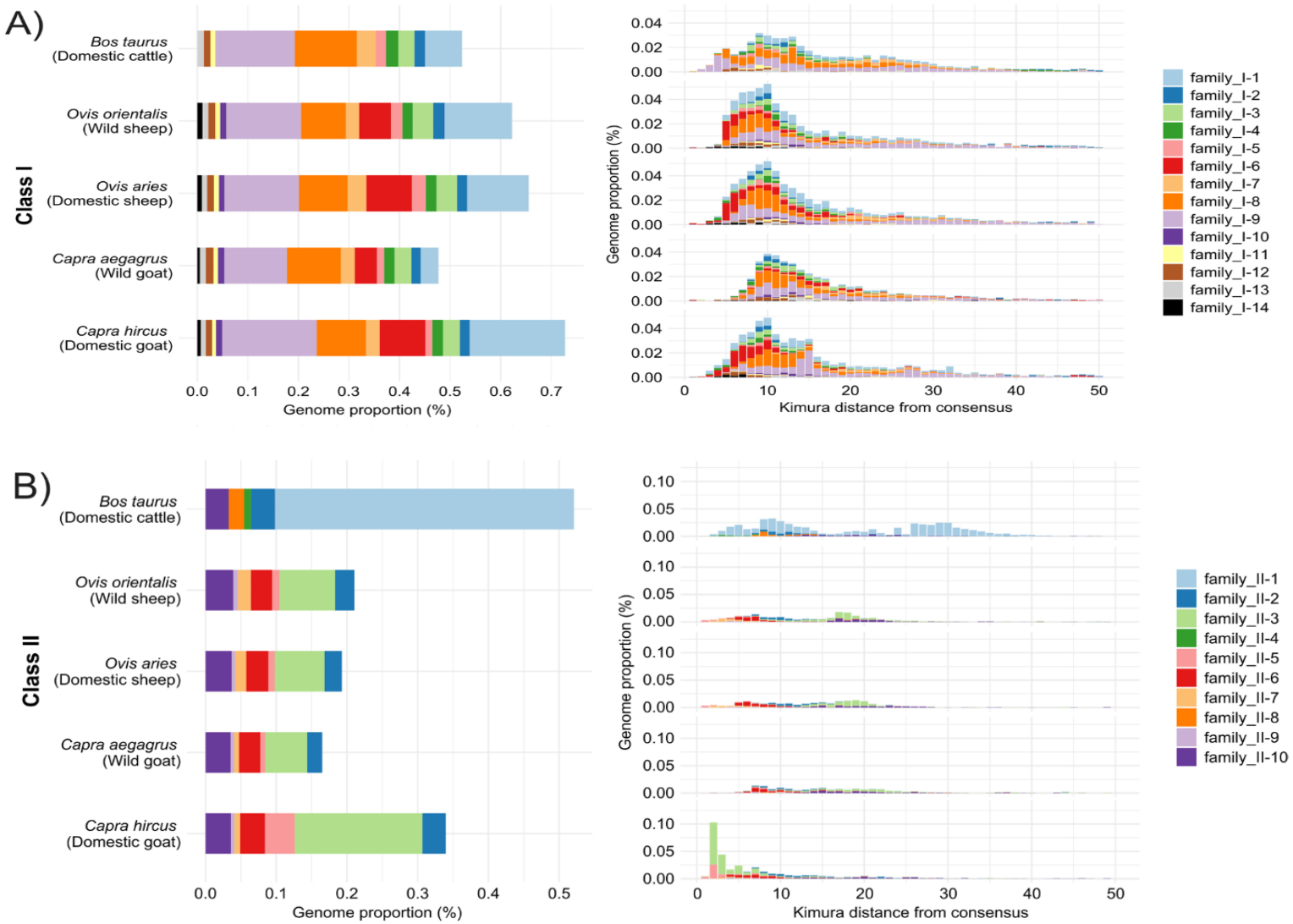
Integration events of ERVs families in ruminants



- Oldest ERV family's integration (I-1, I-5, I-7).
- Most recent ERV family's integration (II-5) and closely related to JSRV and ENTV

Verneret et al mobile DNA

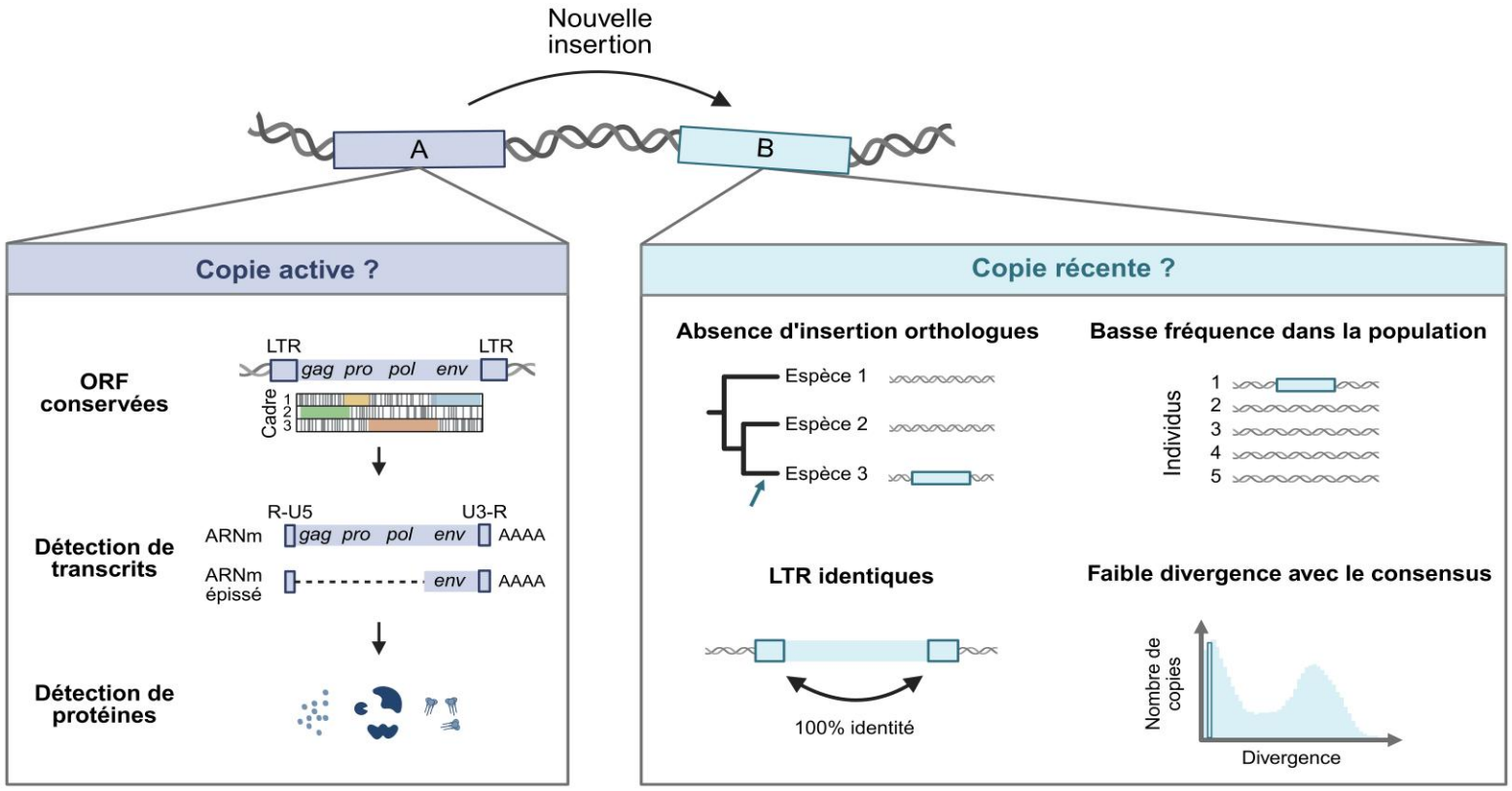
Insertion polymorphism and divergence landscapes of ERV



- 10% kimura distance for ERV1 families.
- ~ 1% kimura distance for ERVK-3 and ERVK-5.

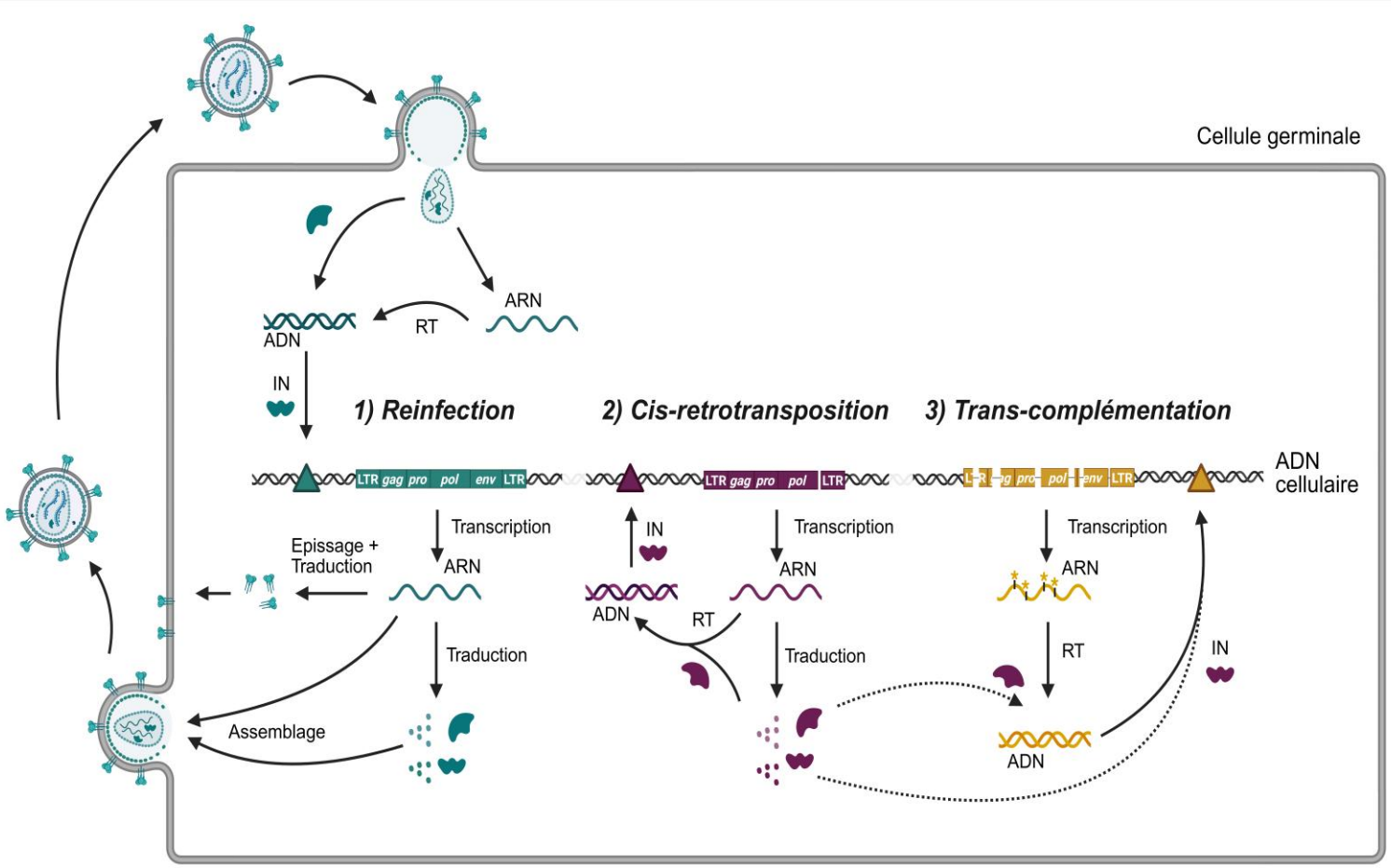
Verneret et al mobile DNA

ERV active and recent copy characteristics



Verneret et al mobile DNA

Propagation of ERV



- ERV with all ORF intact.
- ERV without env.
- ERV with other ORF damage.

Verneret et al

Hypothesis and internship goals

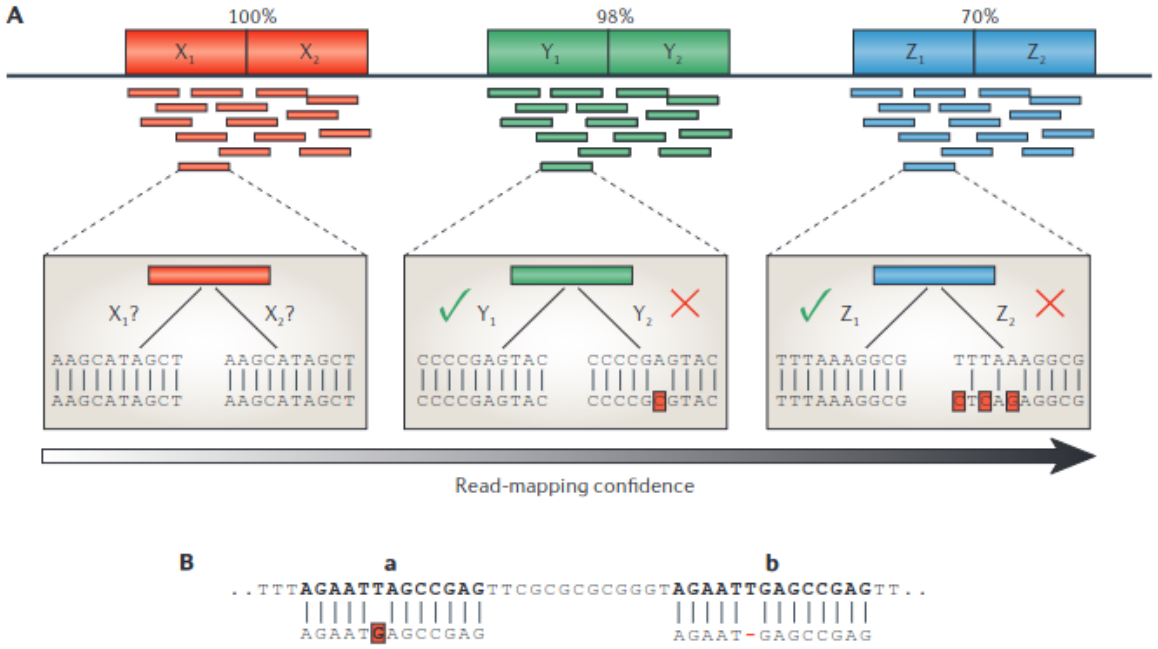
- ERVII-5 and ERVII-3 families could be still active in small ruminant regarding the arguments of this study .

- Collect evidences of ERVII-5 and ERVII-3 are expressed in ....

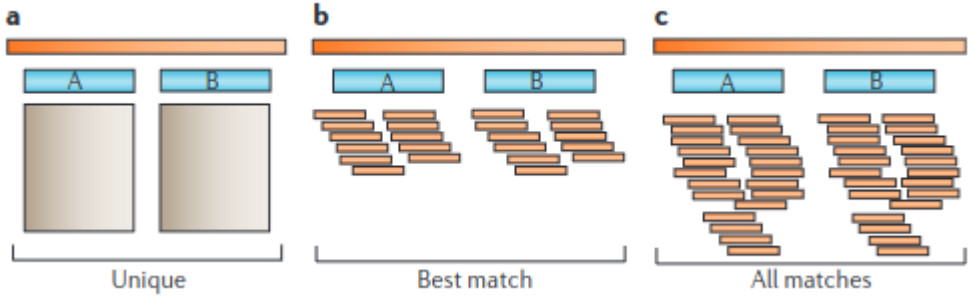
    - Targeted approch

TE mapping difficulties

Differents ways of TE mapping



How tools treat multimap reads ?



- Rigorous choice of tools and parameters are essential.

(Treangen & Salzberg, 2011), NATURE REVIEWS | GENETICS

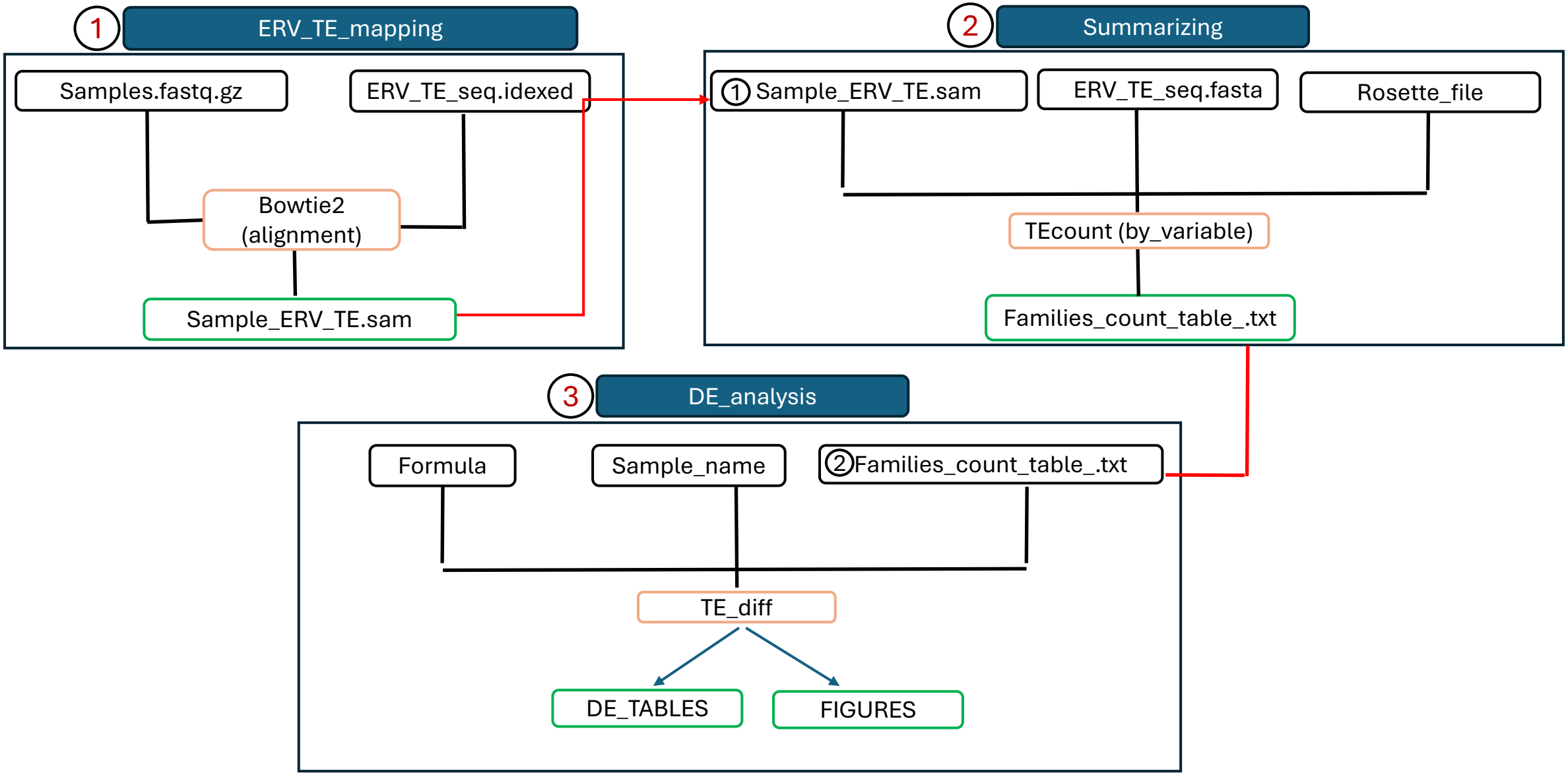
Test samples choice for test

Ovis Aries (PRJNA1017964 , PRJEB19199 *)*	*Capra Hircus* (PRJEB23196)
Uterus X 2 (PRJNA1017964); Lung X 2 (PRJEB19199); Spleen X 2 (PRJEB19199); Liver X 2 (PRJEB19199)	(Uterus, Uterine horn, Ovary, Fallopian tube) X 1; Liver X 2; Spleen X 3; Testis X 3;

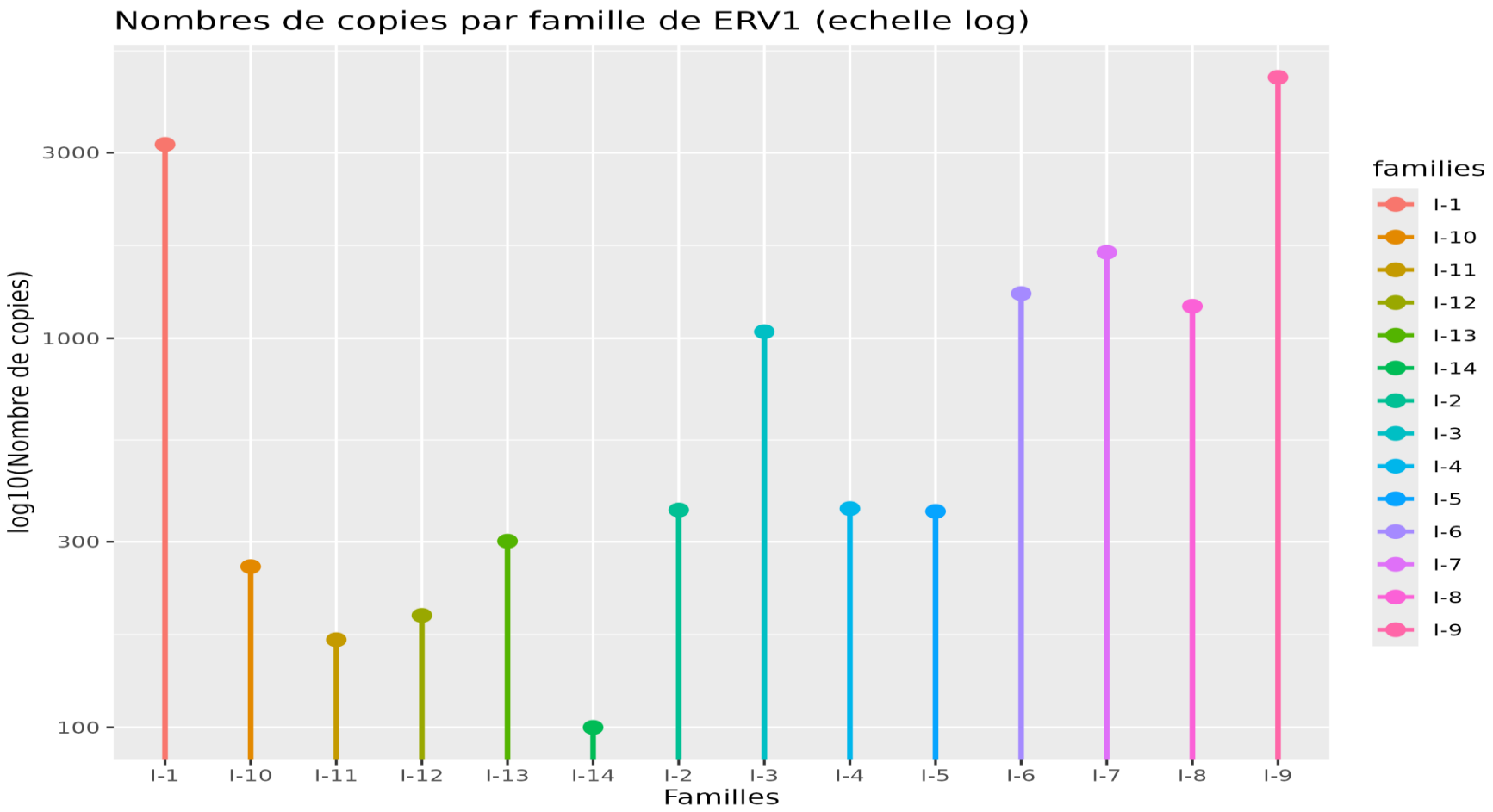
Dev-stage samples choice

Ovis Aries ( )	*Capra Hircus* ()
blastocyst X 3 (PRJNA1126096);	blastocyst X 3 (PRJNA532617, PRJNA543590);
Embryo_2C X 3 (PRJNA1126096);	Embryo_2C X 3 (PRJNA532617, PRJNA543590);
Embryo_4C X 3 (PRJNA1126096);	Embryo_4C X 3 (PRJNA532617, PRJNA543590);
Embryo_8C X 3 (PRJNA1126096);	Embryo_8C X 3 (PRJNA532617, PRJNA543590);
Embryo_16C X 3 (PRJNA1126096);	Embryo_16C X 3 (PRJNA532617, PRJNA543590);
Conceptus_D12 X 4 (PRJNA343223);	
Conceptus_D14 X 4 (PRJNA343223);	
Conceptus_D14 X 4 (PRJNA343223);	
Conceptus_D16 X 4 (PRJNA343223);	
Conceptus_D20 X 4 (PRJNA343223)	

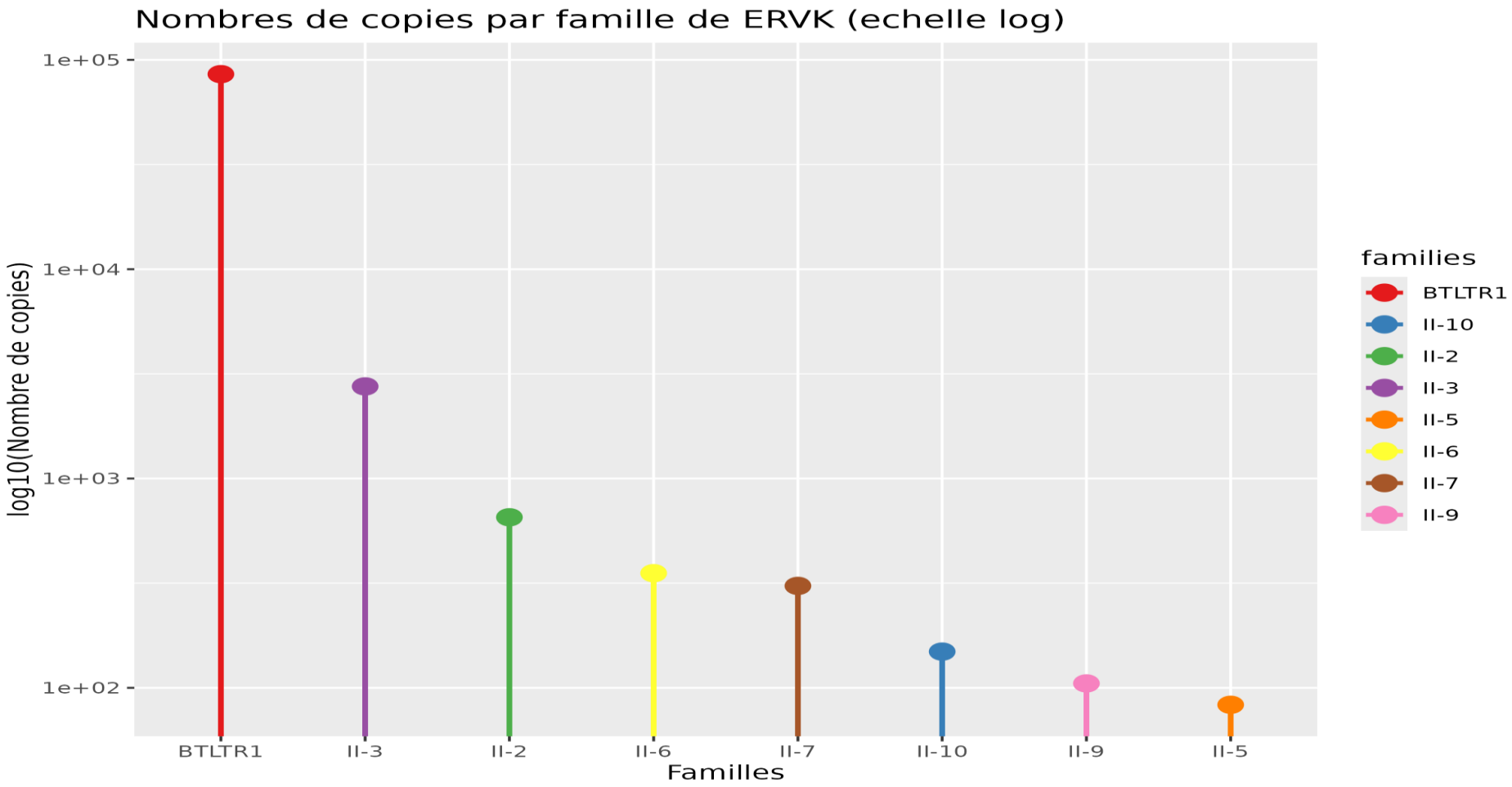
Pipeline


1
ERV_TE_mapping
Samples.fastq.gz
ERV_TE_seq.idexed
Bowtie2
(alignment)
Sample_ERV_TE.sam
2
Summarizing
1 Sample_ERV_TE.sam
ERV_TE_seq.fasta
Rosette_file
TEcount (by_variable)
Families_count_table_.txt
3
DE_analysis
Formula
Sample_name
2 Families_count_table_.txt
TE_diff
DE_TABLES
FIGURES

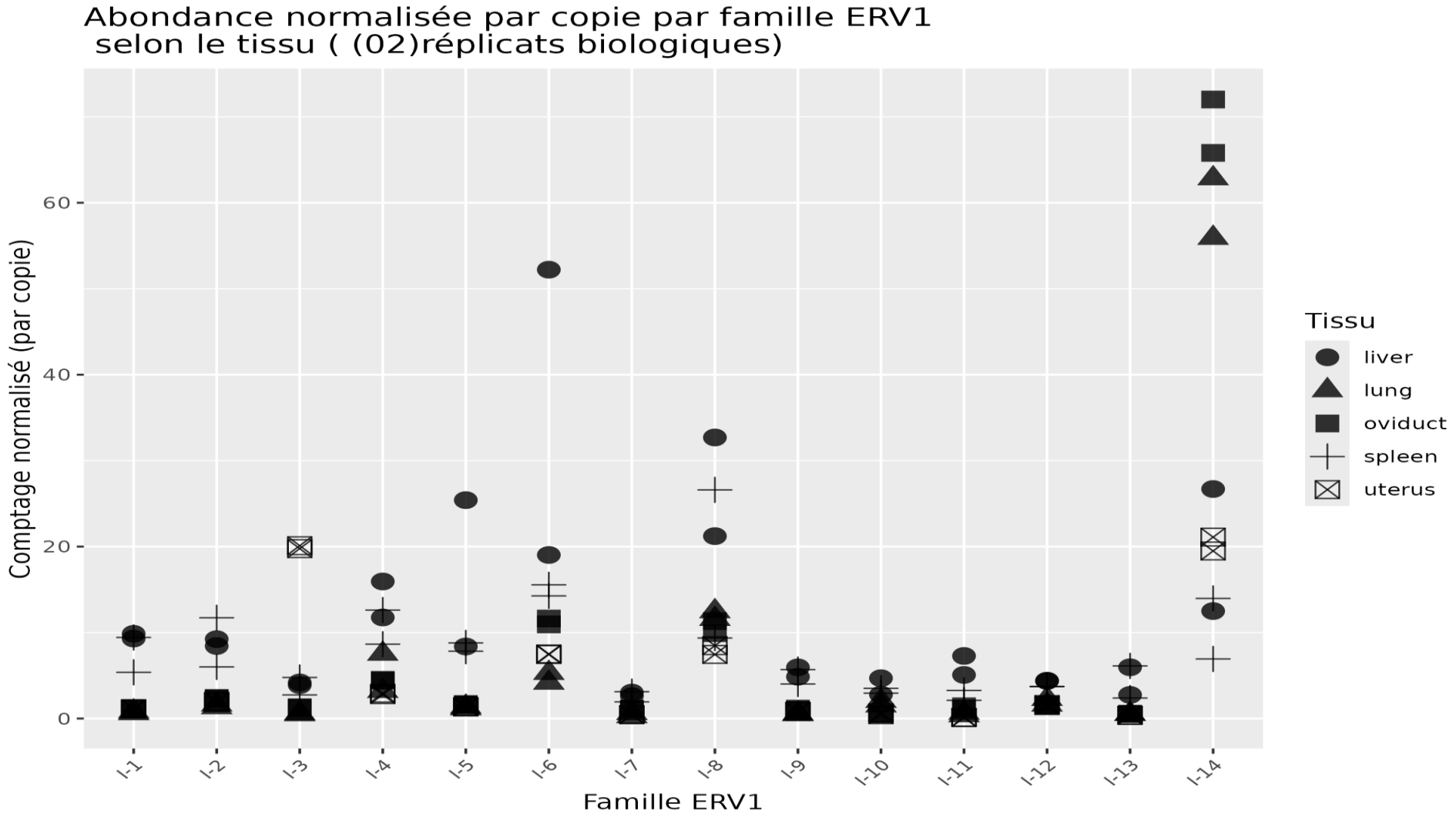
Results
ERV1 copies count in *Ovis aries*



Results
ERVK copies count in *Ovis aries*



Results
ERV1 normalized count in

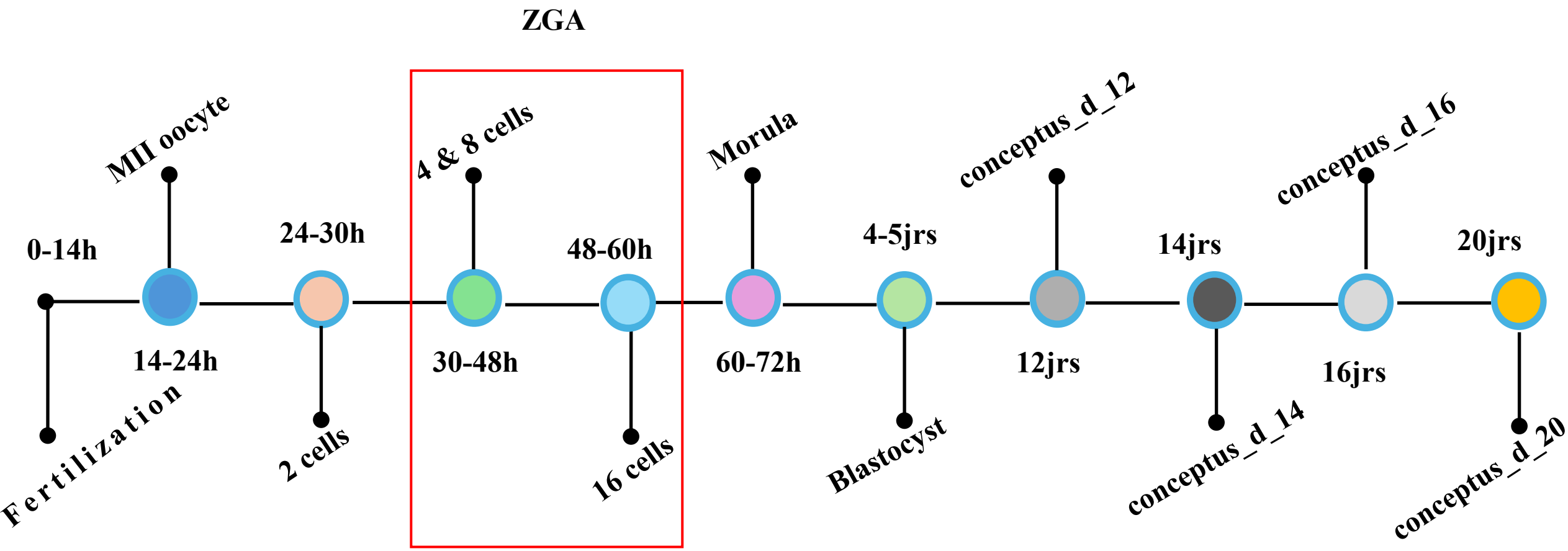


Results
Visualization of ERVK normalized count

Dev-stage sample choice

- Echantillons issus des 4 projets selectionnés sur SRA DB:
  - PRJNA1126096 et PRJNA343223 pour *Ovis aries*.
  - PRJNA532617 et PRJNA543590 pour *Capra hircus*.

Embryo chronologic development in mammals



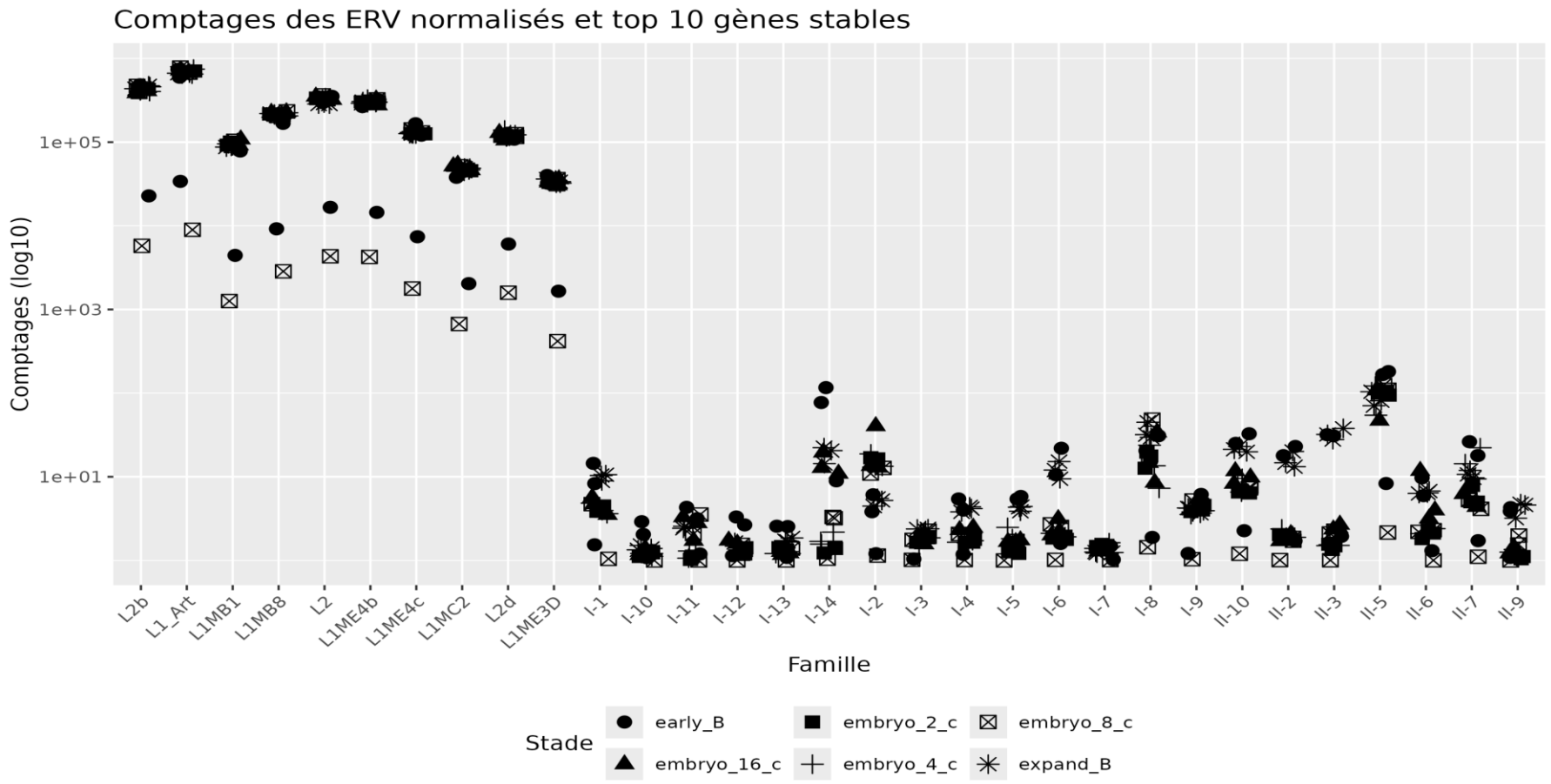
PRJNA1126096

PRJNA1126096: Abnormals samples

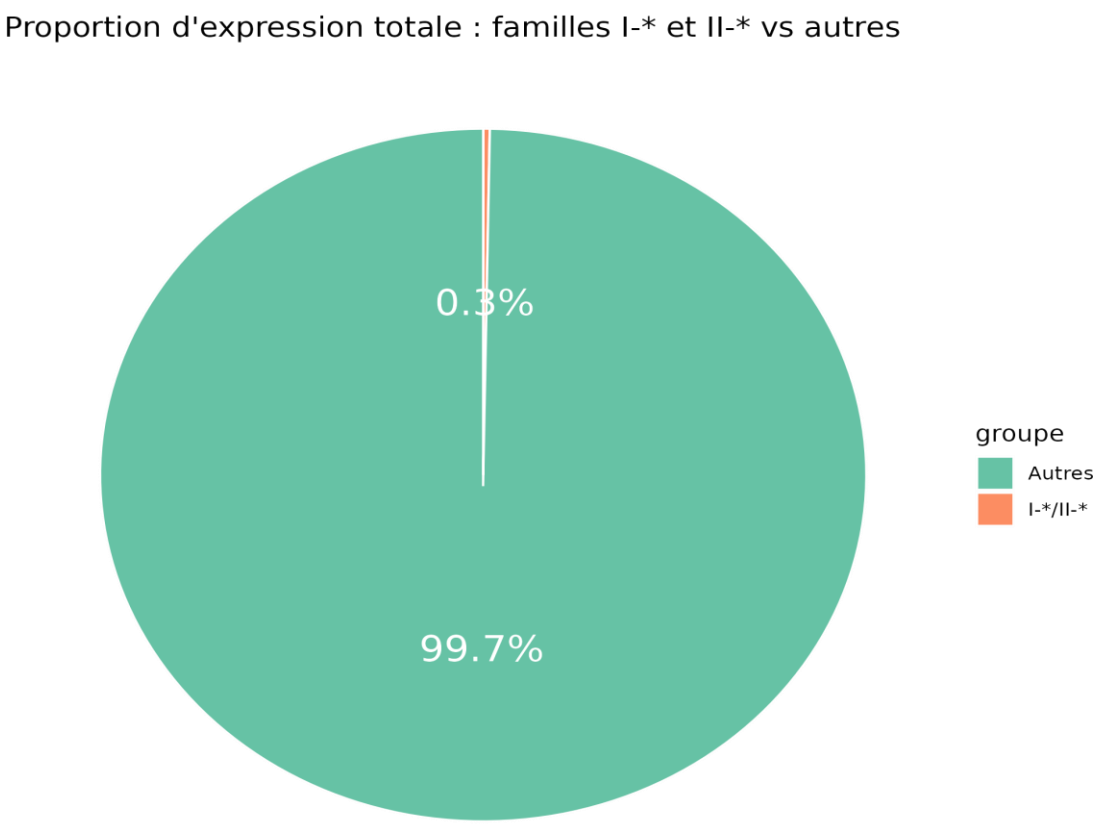
```
early_blastocyst:
total 48G
-rw------- 1 tfousseni bao  50K 10 juil. 10:48 nohup.out
-rw-r--r-- 1 tfousseni bao  24G 10 juil. 15:26 SRR29471657_aln.sam
-rw-r--r-- 1 tfousseni bao  20G 10 juil. 18:18 SRR29471658_aln.sam
-rw-r--r-- 1 tfousseni bao 1,7G 10 juil. 10:44 SRR29471659_1.fastq.gz
-rw-r--r-- 1 tfousseni bao 1,9G 10 juil. 10:47 SRR29471659_2.fastq.gz
-rw-r--r-- 1 tfousseni bao 1,4G 11 juil. 11:24 SRR29471659_aln.sam

expand_blastocyst:
total 67G
-rw------- 1 tfousseni bao 55K 10 juil. 11:00 nohup.out
-rw-r--r-- 1 tfousseni bao 23G 10 juil. 21:16 SRR29471653_aln.sam
-rw-r--r-- 1 tfousseni bao 22G 10 juil. 23:56 SRR29471654_aln.sam
-rw-r--r-- 1 tfousseni bao 22G 11 juil. 02:30 SRR29471655_aln.sam
```

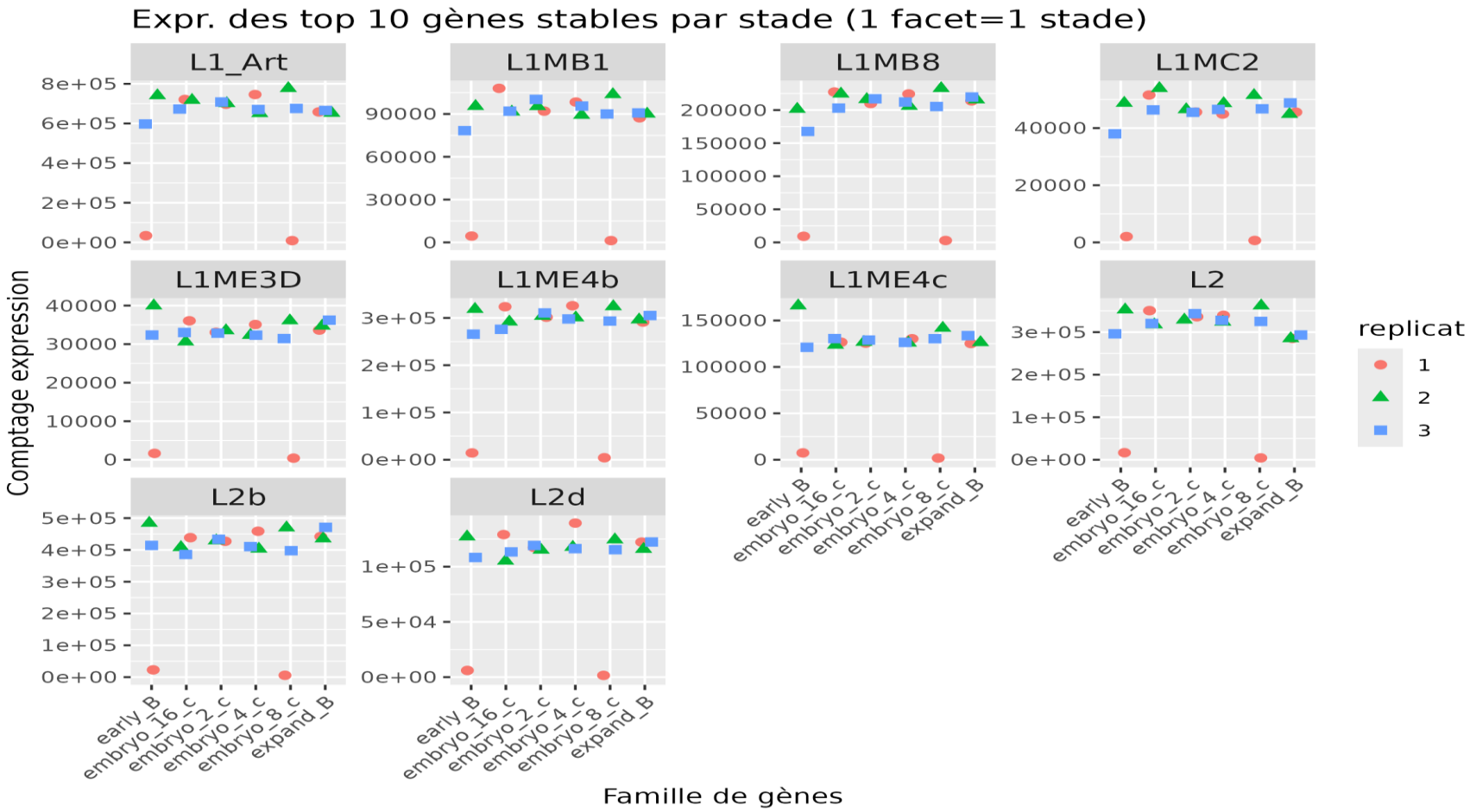
PRJNA1126096: TEcount raw count



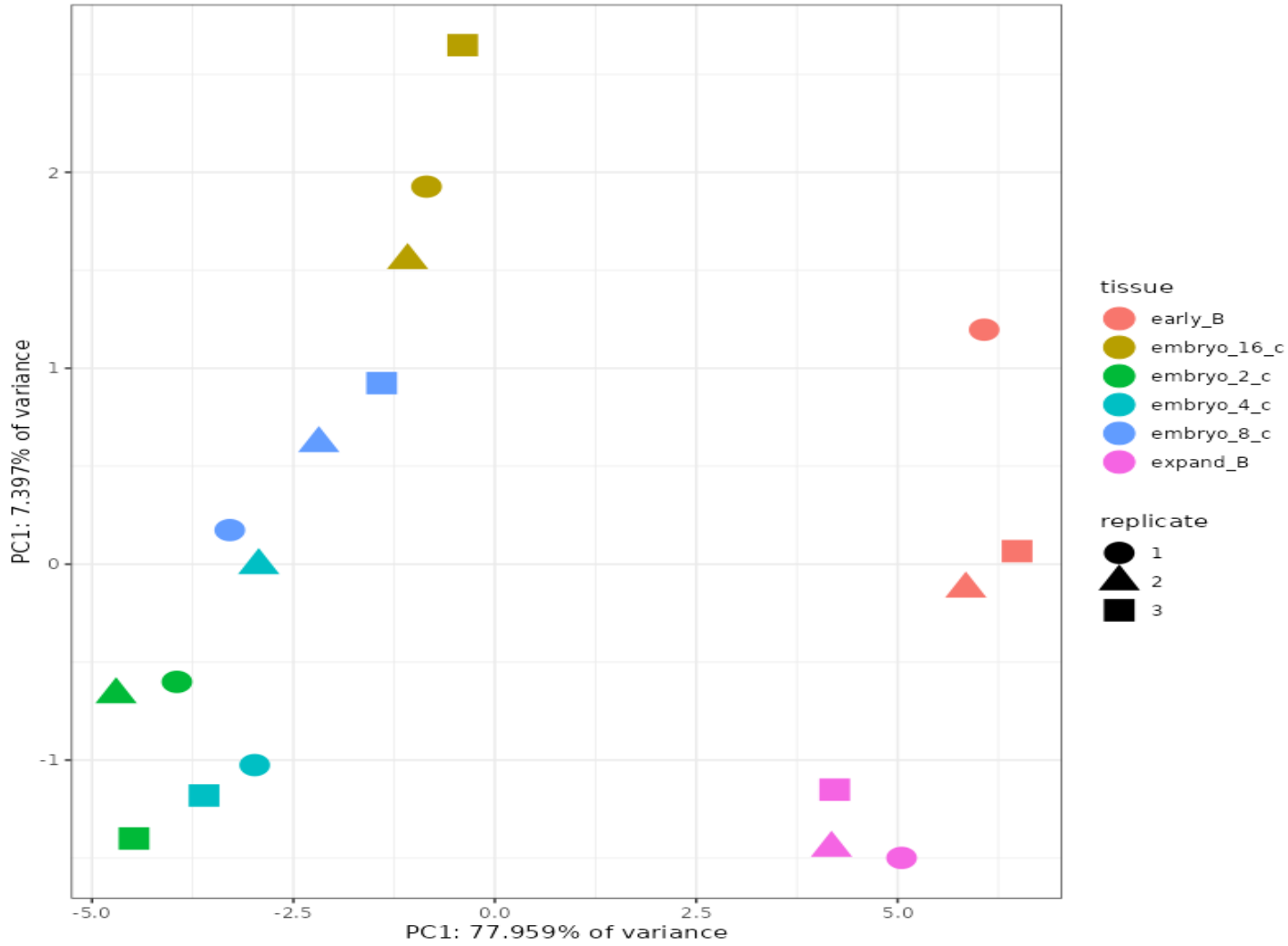
PRJNA1126096: ERV expression proportion in *Ovis aries*



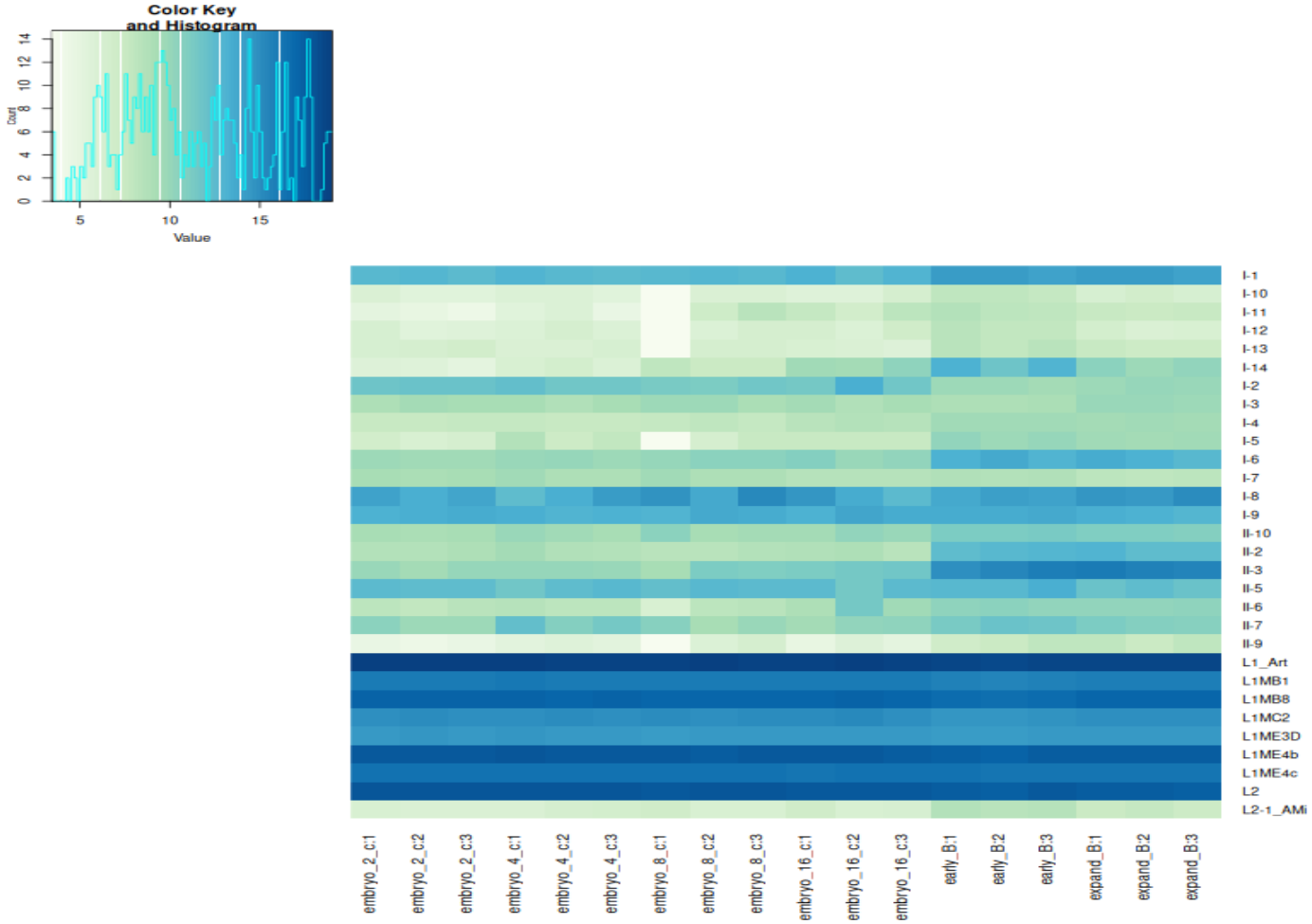
PRJNA1126096: genes with stable expression in blastocyst and embryo



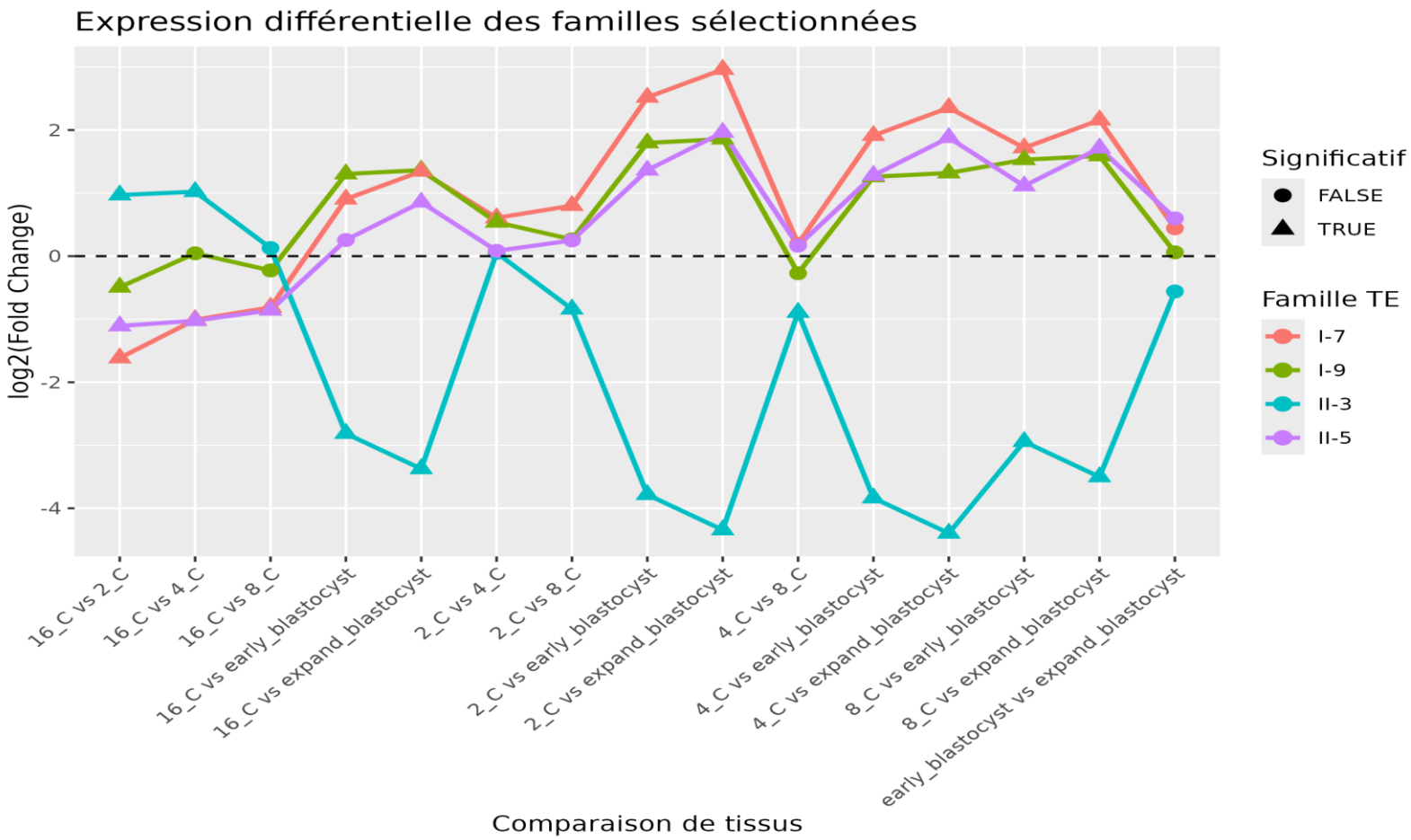
PRJNA1126096: PCA Quality Control



PRJNA1126096: Differential Expression Analysis

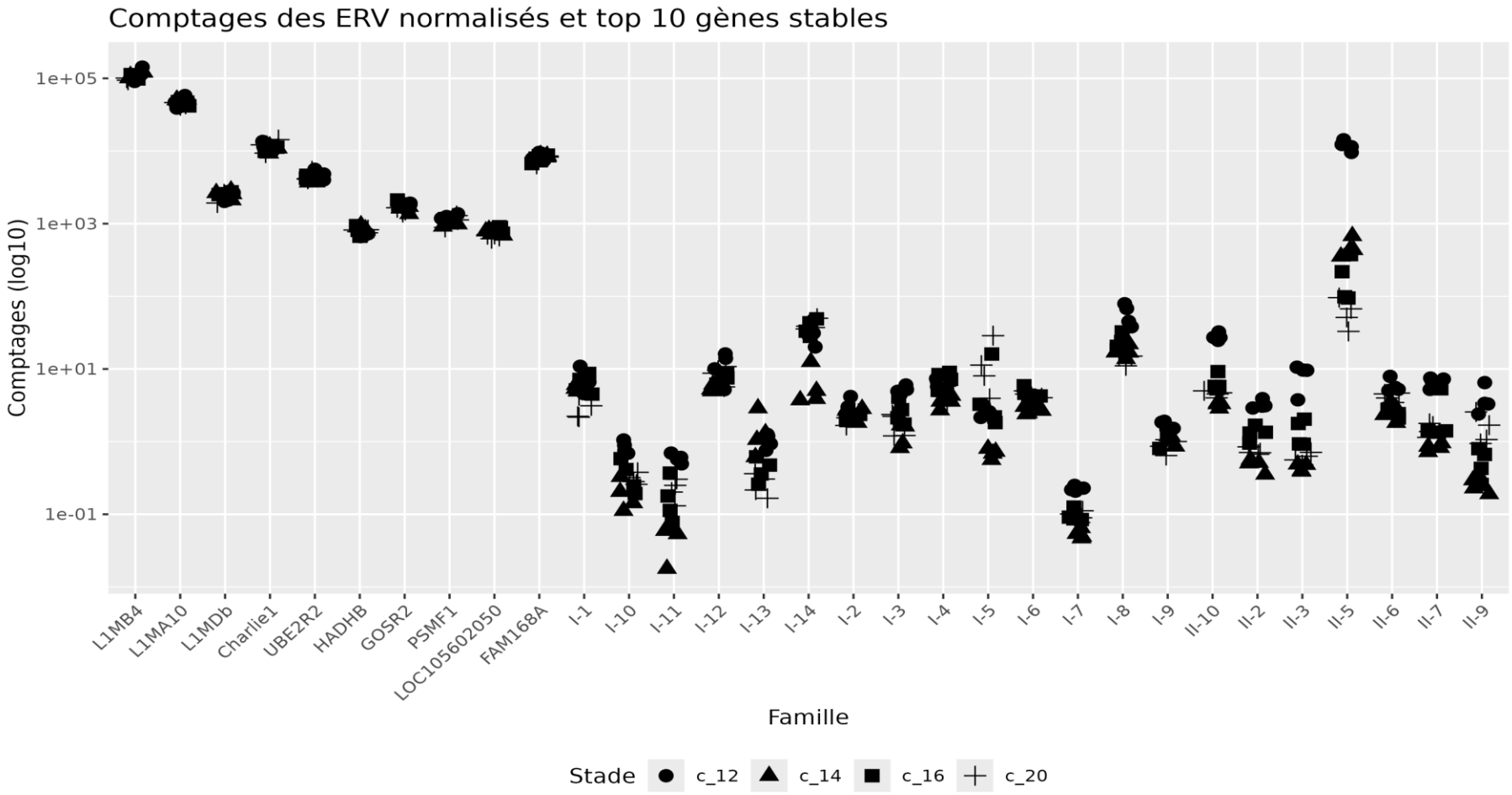


PRJNA1126096: DE of ERV families (I-7, I-9, II-3, II-5)

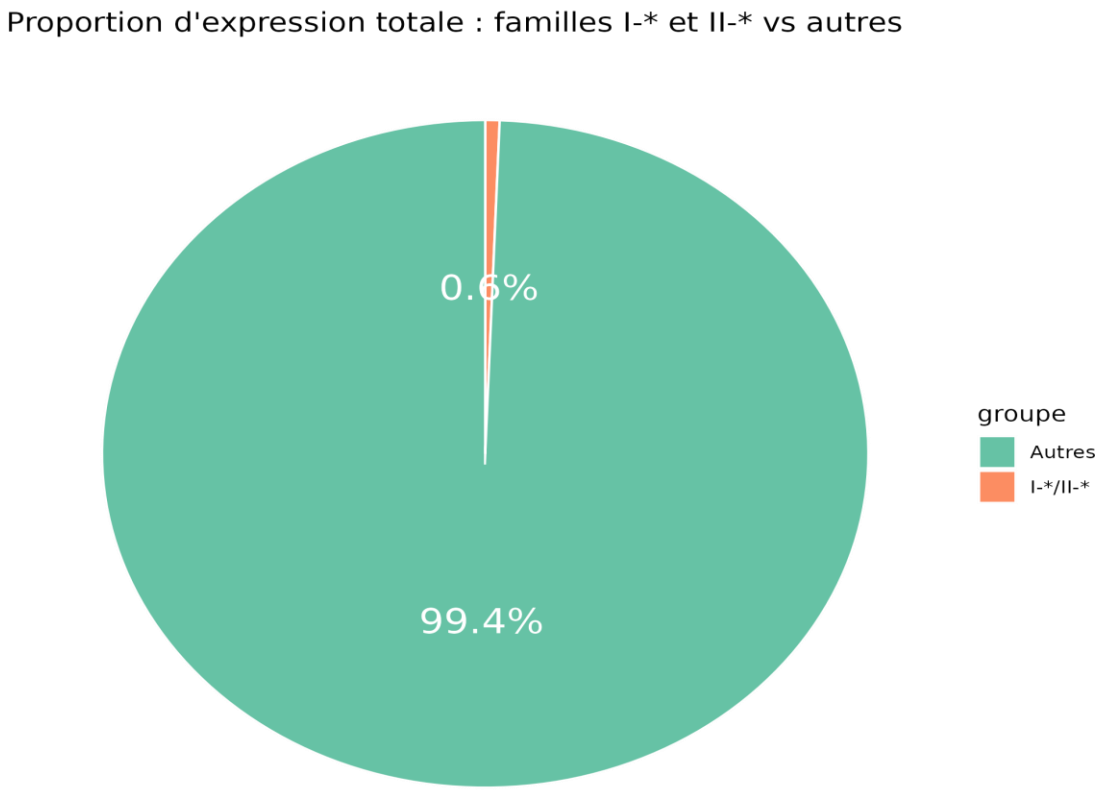


PRJNA343223

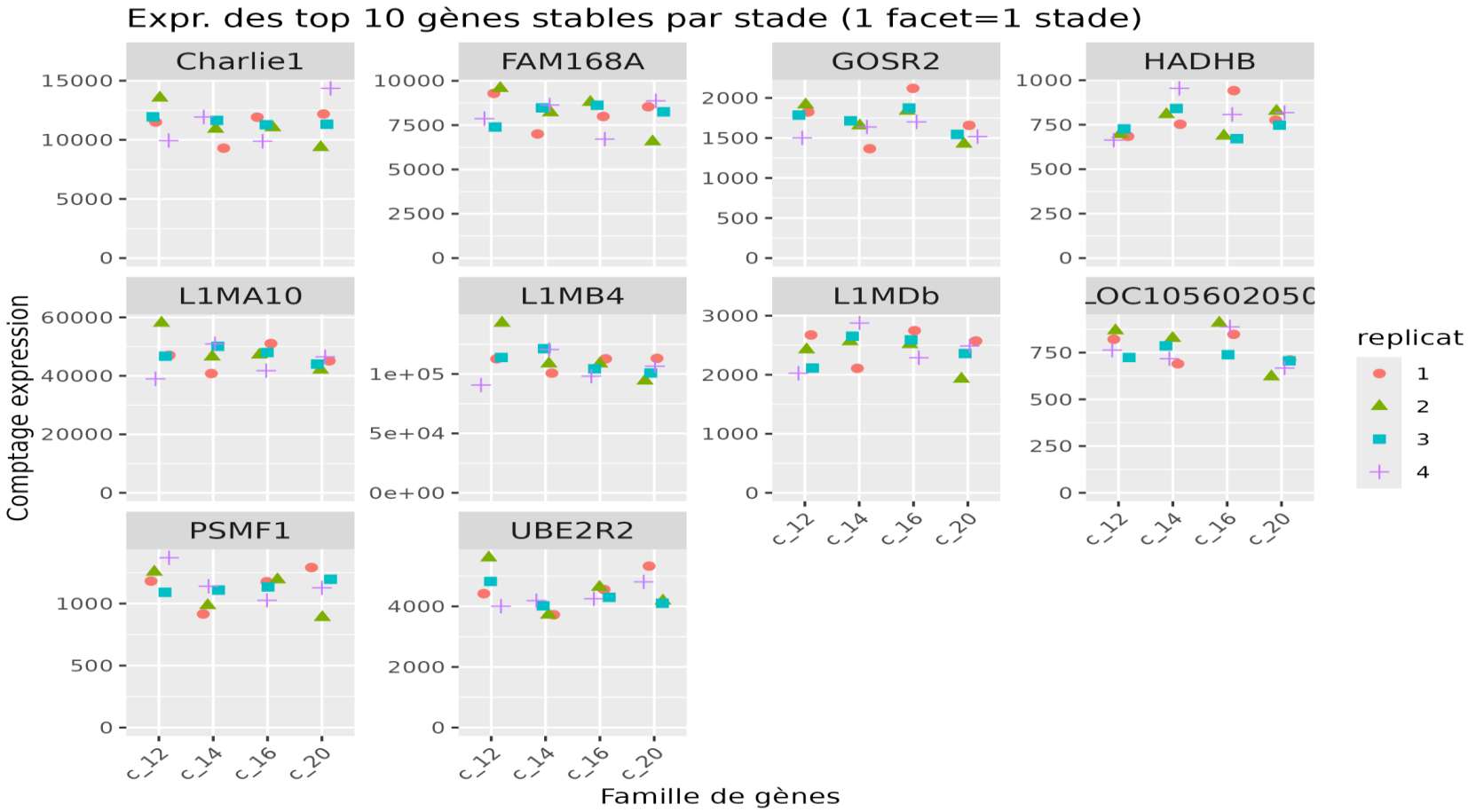
PRJNA343223: Quality control at TEcount level



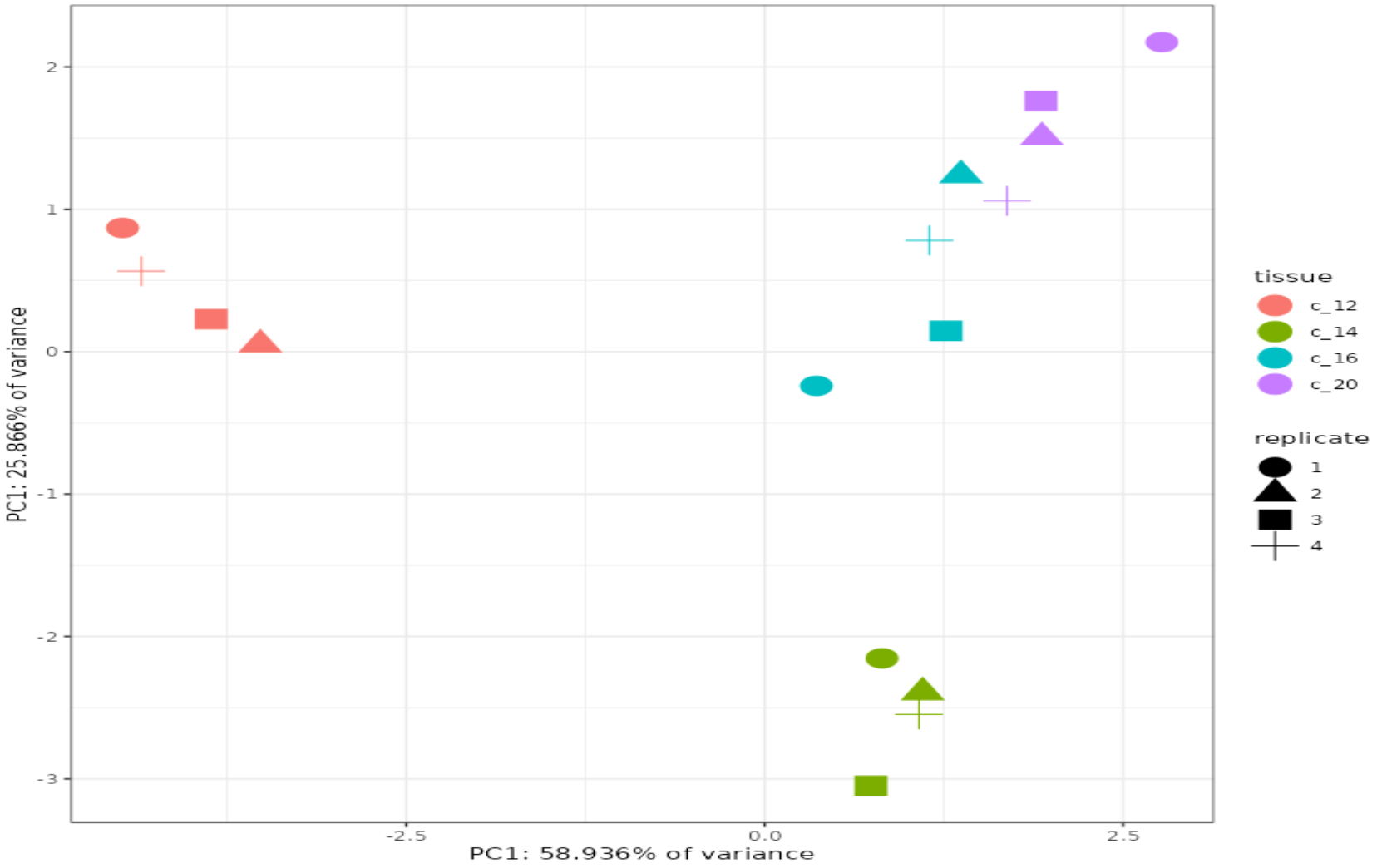
PRJNA343223: ERV expression proportion in *Ovis aries*



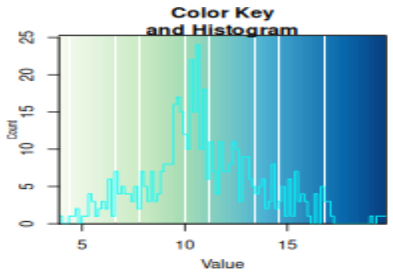
PRJNA343223: genes with stable expression in blastocyst and embryo

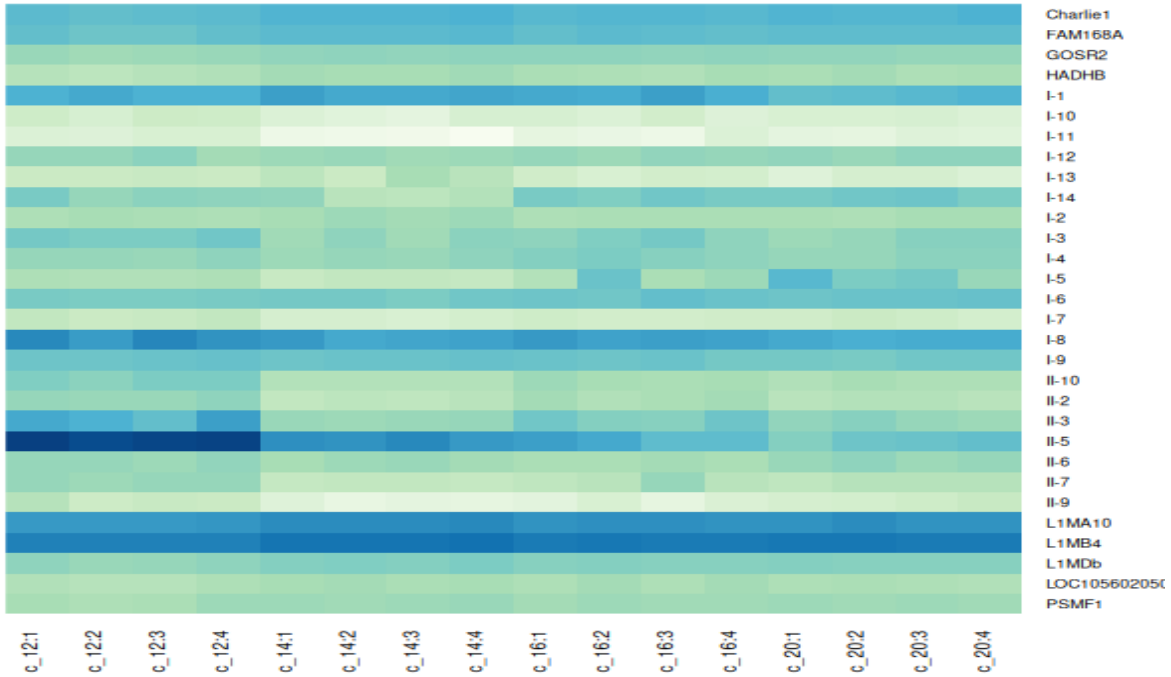


PRJNA343223: PCA Quality Control

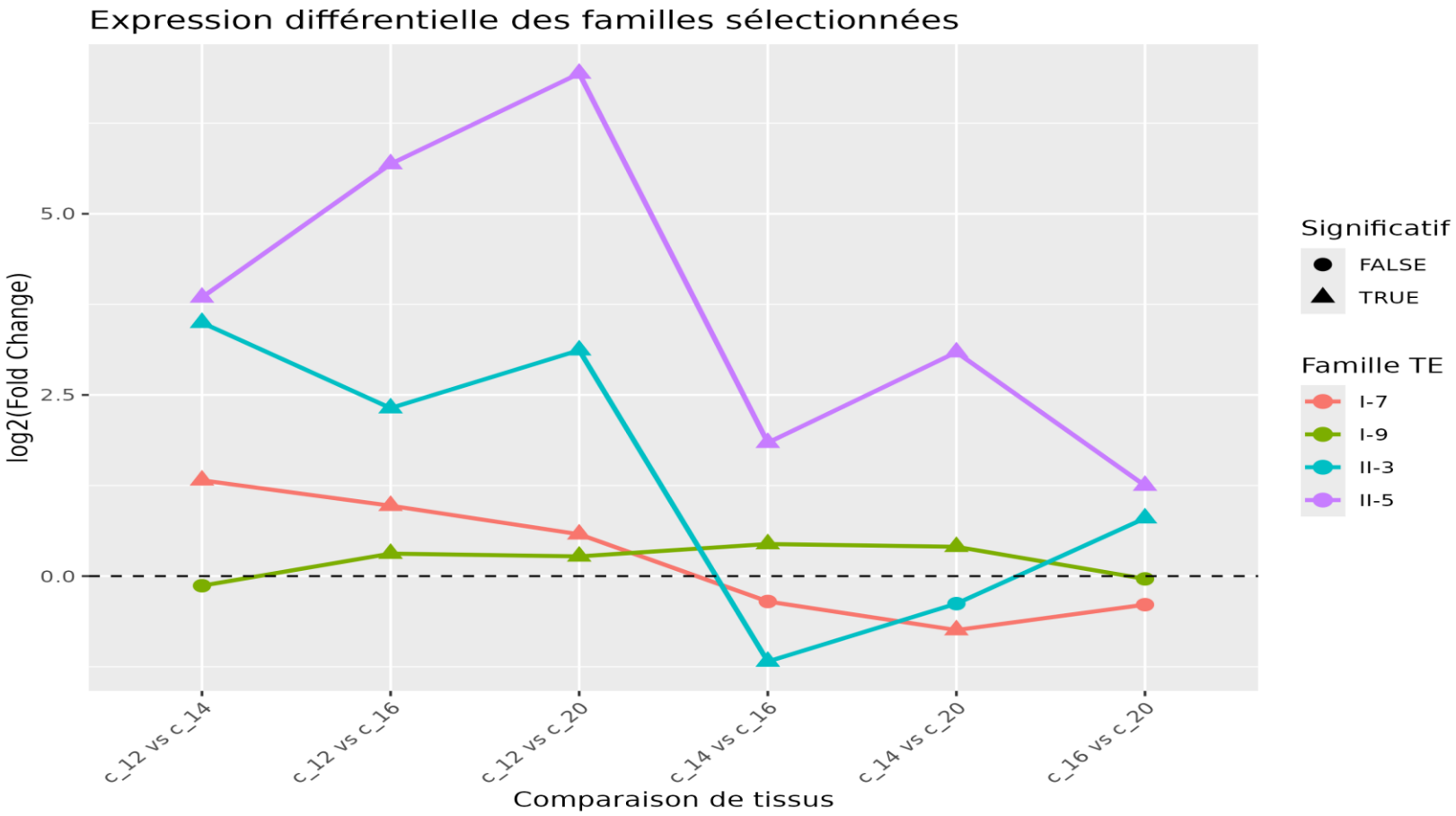


PRJNA343223: Differential Expression Analysis





PRJNA343223: DE of ERV families (I-7, I-9, II-3, II-5)



FIN