# Supplementary material for Part 12 (Feschotte lab analysis)

This page summarizes key findings observed from the analysis of HERV-K polymorphism in the Zimmerome by the Feschotte lab at the University of Utah (Aurélie Kapusta, Cédric Feschotte). Note that we refer Zimmerome as subjectZ in the text below as well as in our presentations.

Schema of data generation and analyses are summarized in the following presentation: **SubjectZ\_HERVK\_schema.pptx**

## Identification of HERV-K copies polymorphic between SubjectZ and the human reference

### a) Method summary

**For more details on the method, see** **SubjectZ\_HERVK\_methods.pdf**

We used the .bam file containing the information of the SubjectZ’s reads mapped to the human reference genome (assembly hg19), to: (i) intersect the coordinates of these reads with coordinates of known polymorphic loci [1,2] and (ii) visualize the mapped reads location by loading the bam file to the UCSC genome browser (thanks: Edward B. Chuong).

Additionally, the same loci were also looked at in the HuRef assembly (Craig Venter’s genome), by blasting the known loci sequences to the HuRef assembly.

### b) Results

We combined the data from intersections and visualization to decipher if HERK-K copies that are absent or present in the reference were absent or present in the SubjectZ’s genome. A total of 132 sites (129 sites listed in ref. [2] and 10 loci listed in ref. [1], 7 loci being in both) were verified.

For five of the 31 loci that do not have a copy of HERV-K in the reference (case 1 from **SubjectZ\_HERVK\_schema.pptx**), we could detect signs that an HERV-K copy was present in SubjectZ’s genome, while at least four of them were absent from the HuRef assembly (one unclear case). We verified that at least some of the SubjectZ reads mapping at that location indeed contain DNA homologous to LTR5\_Hs (which corresponds to the LTR repeats flanking the coding genes of HERV-K). Among the 98 loci that have a copy of HERV-K in the reference (case 2 from **SubjectZ\_HERVK\_schema.pptx**), only one is clearly absent from SubjectZ’s genome (while present in the HuRef assembly).

Four of the HERV-K copies present in the SubjectZ’s genome have been reported in the sequenced Denisovan and/or modern human genomes [3-5] (see attached files below for details).

***Associated data files:***

 **HERV-K\_analysis.xlsx = Details of the loci (excel file)**

 **SubjectZ\_HERV-K\_screenshots.pptx (pptx file)**

***References :***[1] Macfarlane, CM and Badge, RM (2015) Genome-wide amplification of proviral sequences reveals new polymorphic HERV-K(HML-2) proviruses in humans and chimpanzees that are absent from genome assemblies. Retrovirology, Apr 28;12:35.

[2] Wildschutte, JH et al. (2016) Discovery of unfixed endogenous retrovirus insertions in diverse human populations. PNAS, vol.113 no.16.

[3] Agoni L, et al. (2012) Neandertal and Denisovan retroviruses. Curr Biol 22(11):R437–R438.

[4] Lee E, et al. (2012) Landscape of somatic retrotransposition in human cancers. Science 337(6097):967–971.

[5] Lee A, et al. (2014) Novel Denisovan and Neanderthal retroviruses. J Virol 88(21):12907–12909.

[6] Marchi E, et al. (2014) Unfixed endogenous retroviral insertions in the human population. J Virol 88(17):9529–9537