# Comparison of TEtranscripts and the new pipeline

Contents

[Comparison of TEtranscripts and the new pipeline 1](#_Toc532992305)

[Comparison of TEtranscripts and the new pipeline 1](#_Toc532992306)

[1. Quantification of read count using TEtranscripts 1](#_Toc532992307)

[2. Capture of expressed TE families from TEtranscripts data 2](#_Toc532992308)

[3. Comparison of the expressed TE families based on TEtranscripts and the new pipeline 2](#_Toc532992309)

# Comparison of TEtranscripts and the new pipeline

## Quantification of read count using TEtranscripts

#!/bin/bash

#TEtranscript reverse

TEtranscripts -c 01\_k100MmNamesorted.bam 02\_k100MmNamesorted.bam **\**

03\_k100MmNamesorted.bam 04\_k100MmNamesorted.bam 05\_k100MmNamesorted.bam **\**

06\_k100MmNamesorted.bam 07\_k100MmNamesorted.bam 08\_k100MmNamesorted.bam **\**

09\_k100MmNamesorted.bam 10\_k100MmNamesorted.bam 11\_k100MmNamesorted.bam **\**

12\_k100MmNamesorted.bam 13\_k100MmNamesorted.bam 14\_k100MmNamesorted.bam **\**

15\_k100MmNamesorted.bam **\**

-t 16\_k100MmNamesorted.bam 17\_k100MmNamesorted.bam 18\_k100MmNamesorted.bam **\**

19\_k100MmNamesorted.bam 20\_k100MmNamesorted.bam 21\_k100MmNamesorted.bam **\**

22\_k100MmNamesorted.bam 23\_k100MmNamesorted.bam 24\_k100MmNamesorted.bam **\**

25\_k100MmNamesorted.bam 26\_k100MmNamesorted.bam 27\_k100MmNamesorted.bam **\**

28\_k100MmNamesorted.bam 29\_k100MmNamesorted.bam 30\_k100MmNamesorted.bam **\**

31\_k100MmNamesorted.bam 32\_k100MmNamesorted.bam 33\_k100MmNamesorted.bam **\**

34\_k100MmNamesorted.bam 35\_k100MmNamesorted.bam 36\_k100MmNamesorted.bam **\**

37\_k100MmNamesorted.bam 38\_k100MmNamesorted.bam 39\_k100MmNamesorted.bam **\**

**--**GTF '../Reference/Curated/V2.1\_iso1\_exon.gtf' **\**

**--**TE '../Reference/Curated/AllTEsExpanded\_curated\_V2.gtf' **\**

**--**mode multi **\**

**--**stranded reverse **\**

**--**project **./**TEtranscripts\_k100MmMultiReverse

## Capture of expressed TE families from TEtranscripts data

#R code

RE **=** read.table**(**'TEtranscripts\_k100MmMultiReverse.cntTable', header**=**T, row.names**=**1**)**

RE **=** RE**[**,order**(**names**(**RE**))]**

RE**=**RE**[!**grepl**(**"\_\_", row.names**(**RE**))**,**]**

RE**=**RE**[!**grepl**(**"VIT\_", row.names**(**RE**))**,**]**

RE2**=**RE**[**,1**:**3**]** #ctrl

RE2**$**mean**=**rowMeans**(**RE2**[**,1**:**3**])**

RE2**=**subset**(**RE2, RE2**$**mean**>**10**)**

RE2**$**id**=**rownames**(**RE2**)**

RE2**=**RE2**[**,c**(**5,1**:**4**)]**

write.table**(**RE2, "TEtranscripts\_k100MmMultiReverse\_exprCtrl10.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Comparison of the expressed TE families based on TEtranscripts and the new pipeline

#R code

TEtranscript**=**read.table**(**'TEtranscripts\_k100MmMultiReverse\_exprCtrl10.txt', header**=**T, sep**=**"\t"**)**

newpipeline**=**read.table**(**'Candidates\_REfamily\_tag\_ctrl\_new.txt', header**=**T, sep**=**"\t"**)**

newpipeline**=**subset**(**newpipeline, candidate**>**0**)**

#generate Venn Diagram

library**(**VennDiagram**)**

v**<-**venn.diagram**(**list**(**"TEtranscript"**=**TEtranscript**[**,1**]**, "newpipeline"**=**newpipeline**[**,1**])**, filename**=NULL)**

grid.newpage**()**

pdf**(**file **=** "VennDiagram\_NewPipelineVsTEtranscripts\_ctrl.pdf", width**=**7, height**=**7**)**

pushViewport**(**viewport**(**width**=**unit**(**0.75, "npc"**)**, height **=** unit**(**0.75, "npc"**)))**

grid.draw**(**v**)**

dev.off**()**

#check on the TE family uniquely found by TEtranscripts of the pipeline

diff.id**=(!**TEtranscript**$**id %in% newpipeline**$**id**)**

TEtUniq**=**TEtranscript**[**diff.id,**]**

diff.id**=(!**newpipeline**$**id %in% TEtranscript**$**id**)**

NPUniq**=**newpipeline**[**diff.id,**]**

#Expressed TE family uniquely found based on TEtranscripts results:

#Copia-12

#Copia-15

#Copia-38

#Copia-45

#Copia-58

#Copia-7

#Copia-9

#Gypsy-10

#Harbinger-3N3

#MULE-MuDR-13

#Expressed TE family uniquely found based on the new pipeline:

#Copia-69

#Copia-84

#Copia-93

#CACTA-4

#hAT-Vinesleeper-2

#MULE-MUGvine-1

#MULE-MUGvine-2

#Check the expression range of the TE family uniquely found by TEtranscripts

count**=**read.table**(**"../BedCov\_OverlapBP\_sense\_count.txt", header**=**T, sep**=**"\t"**)**

depth**=**read.table**(**"../BedCov\_OverlapBP\_sense\_depth.txt", header**=**T, sep**=**"\t"**)**

count1**=**count**[**,1**:**9**]**

depth1**=**depth**[**,1**:**9**]**

count1**$**countMean**=**rowMeans**(**count1**[**,7**:**9**])**

depth1**$**depthMean**=**rowMeans**(**depth1**[**,7**:**9**])**

df**=**cbind**(**count1**[**,c**(**1**:**6,10**)]**, depthMean**=**depth1**$**depthMean**)**

dfNonZero**=**subset**(**df, df**$**countMean**>**0**)**

dfNonZero**<-**dfNonZero**[**order**(**dfNonZero**$**TEm**)**,**]**

TElist**=**paste**(**TEtUniq**$**id, "\_", sep**=**""**)**

data**=**dfNonZero**[**grepl**(**paste**(**TElist, collapse**=**"|"**)**, dfNonZero**$**TEm**)**,**]**

data**<-**data**[**order**(**data**$**id**)**,**]**

library**(**ggplot2**)**

ggplot**(**data, aes **(**log2**(**countMean**)**, log2**(**depthMean**)))+**

theme\_bw**()+**

geom\_point**(**colour**=**"goldenrod2", alpha **=** 0.5, size**=**5**)+**

geom\_vline**(**xintercept **=** log2**(**10**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_vline**(**xintercept **=** log2**(**8**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_vline**(**xintercept **=** log2**(**6**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_vline**(**xintercept **=** log2**(**4**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_vline**(**xintercept **=** log2**(**2**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_hline**(**yintercept **=** log2**(**5**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_hline**(**yintercept **=** log2**(**4**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_hline**(**yintercept **=** log2**(**3**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

geom\_hline**(**yintercept **=** log2**(**2**)**, linetype **=** 2, size**=**1.5, colour **=** "gray35"**)+**

xlim**(NA**, 15**)+**

ylim**(NA**, 10**)**

dev.copy**(**png,"TE\_TEtranscripts\_unique\_ctrl.png", width **=**1200, height **=** 900**)**

dev.off**()**