Step-by-step tutorial for TransTag library analysis:

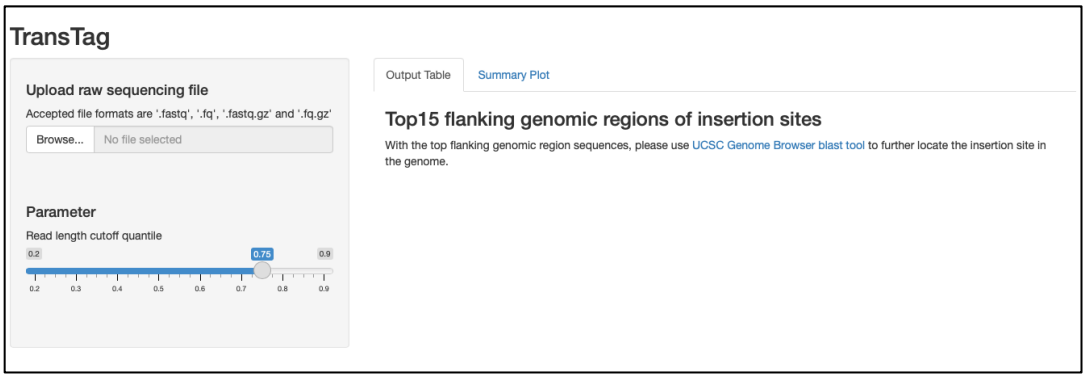
Alignment-free analysis using R Shiny App

For alignment-free sequencing analysis, we provide a Shiny App interface to identify transgene insertion sites (https://menglab.shinyapps.io/transtag_alignmentfree/). Briefly, the Shiny App identifies chimeric reads from raw sequencing data, removes Tn5 adapter and Tol2 sequences to extract the genomic sequences flanking transgene insertion site, and displays the top fifteen most abundant flanking region sequences. The enriched flanking sequences can then be mapped to the zebrafish genome using any standard online tool, such as UCSC BLAT (https://genome.ucsc.edu) or NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), allowing users to assign genomic coordinates of transgenes.

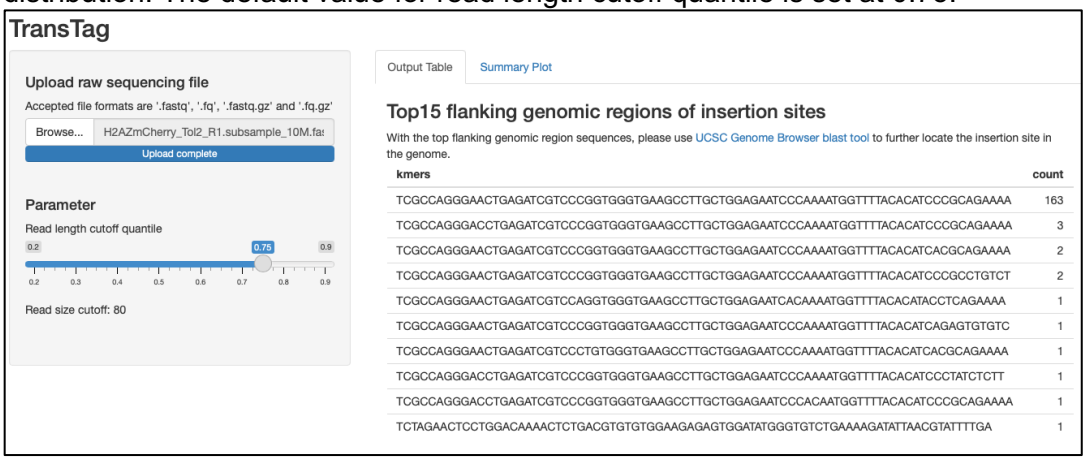
- **Using the online TransTag Shiny app interface** (for smaller size input files (<1 GB))

  a. Open alignment-free R Shiny App website in a web browser using the following link https://menglab.shinyapps.io/transtag_alignmentfree/

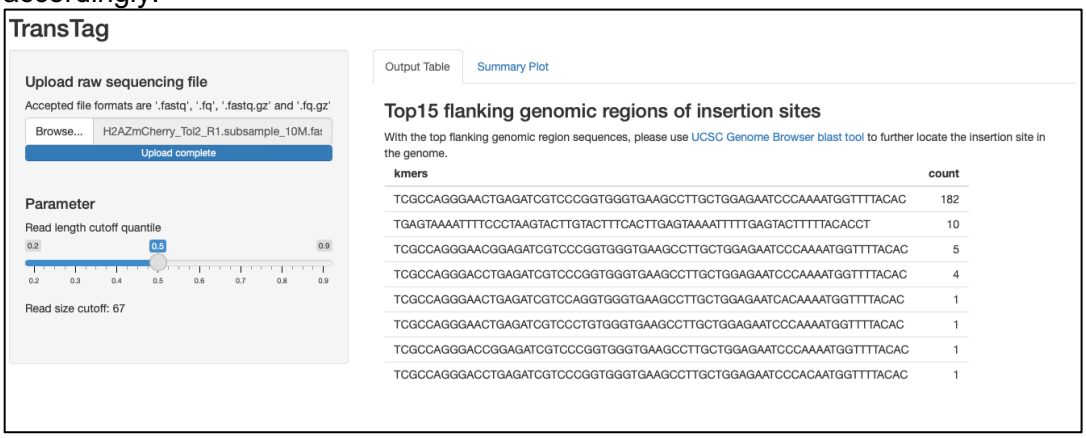
  Below is the user interface of the TransTag Shiny App:



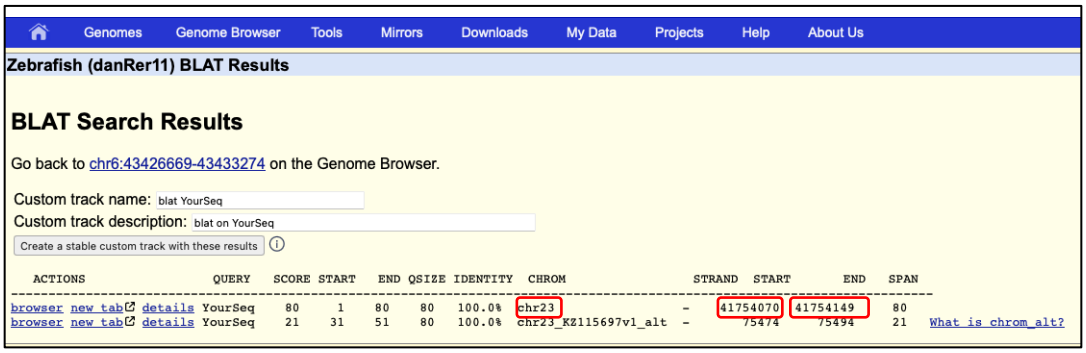
  b. Upload raw sequencing data for processing. It will show the top15 abundant genomic sequences flanking Tol2 insertion site in the “Output Table” tab, and size distribution of the chimeric reads with flanking genomic sequences in the “Summary Plot” tab. You may choose to change the read length cutoff quantile between 0.2 to 0.9 based on the size distribution. The default value for read length cutoff quantile is set at 0.75.



c. By changing the read length cutoff quantile between 0.2 to 0.9, the TransTag Shiny app will display flanking genomic regions with different lengths and count numbers accordingly.

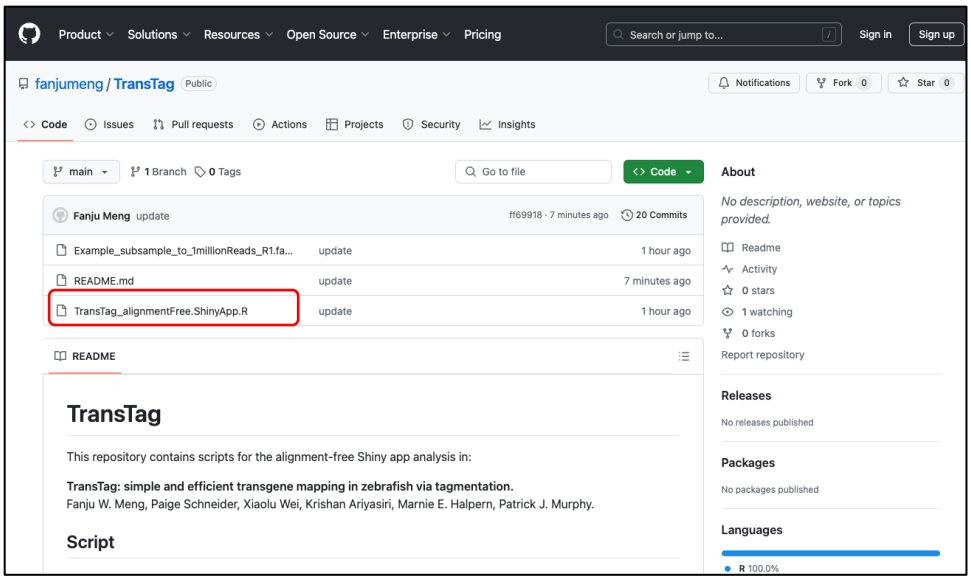


d. To identify the genomic coordinates of transgene insertion site(s), use any standard online tool, such as UCSC BLAT (https://genome.ucsc.edu) or NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

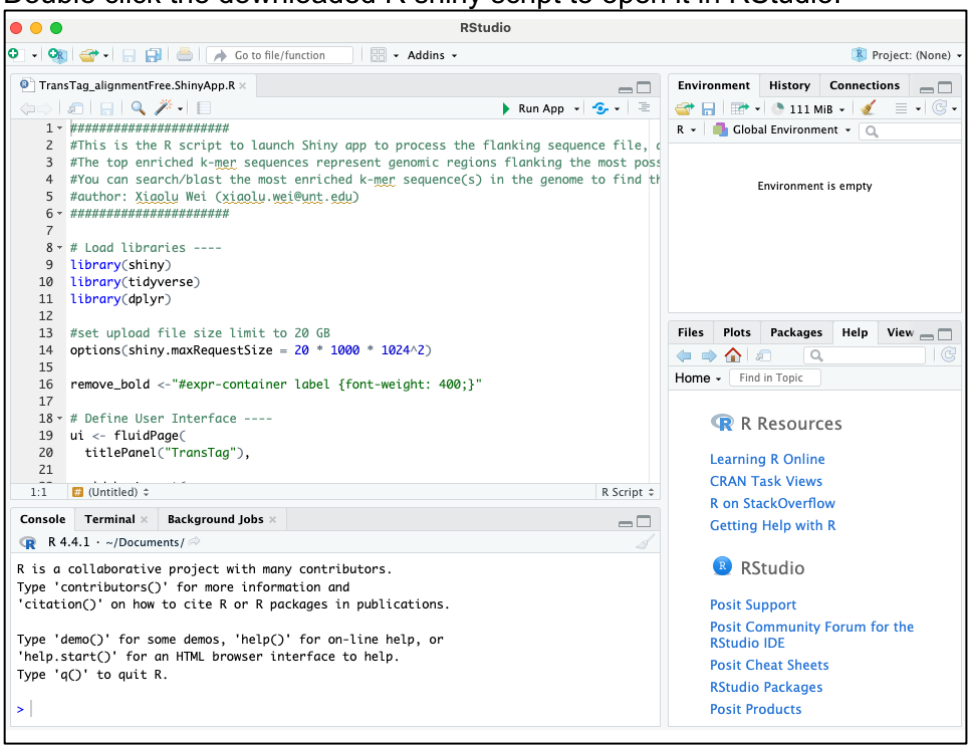


- **Running the shiny script on local computers** (for any size input files)

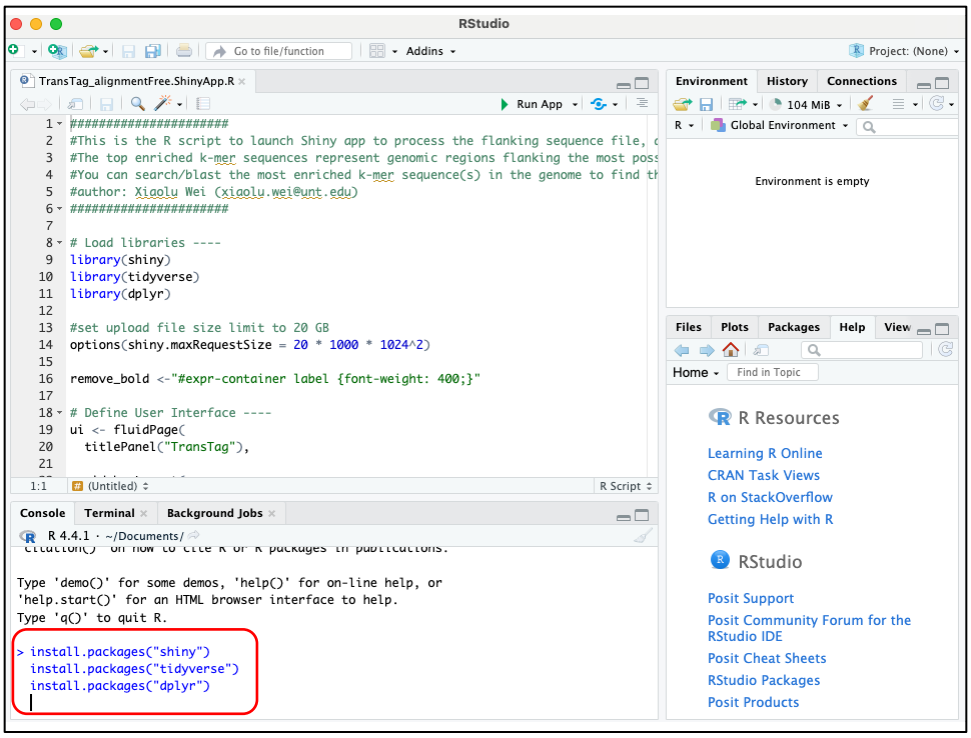
a. Download and install R and RStudio on local computer following instruction: https://posit.co/download/rstudio-desktop/
b. Download alignment-free R Shiny app code “TransTag_alignmentFree.ShinyApp.R” from GitHub at https://github.com/fanjumeng/TransTag.



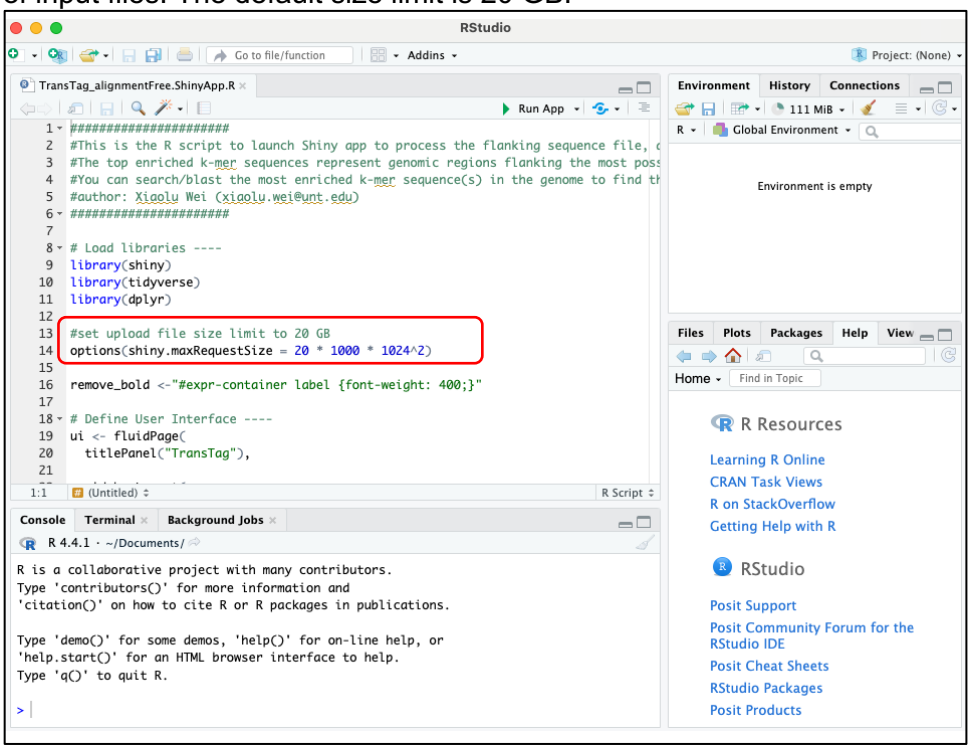
c. Double click the downloaded R shiny script to open it in RStudio.



d. Install required packages. Type the following code in RStudio Console and enter to run:
install.packages("shiny")
install.packages("tidyverse")
install.packages("dplyr")



e. Modify options (shiny.maxRequestSize =) in the R shiny script to increase the size limit of input files. The default size limit is 20 GB.



f. Click “**Run App**” to launch the Shiny app. The same interface as the TransTag Shiny App website will show up. Please follow the same steps mentioned in the above “**Using the online TransTag Shiny app interface**” section to conduct analysis.