

## 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/](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))

- Open alignment-free R Shiny App website in a web browser using the following link [https://menglab.shinyapps.io/transtag\\_alignmentfree/](https://menglab.shinyapps.io/transtag_alignmentfree/)

Below is the user interface of the TransTag Shiny App:

The screenshot shows the 'TransTag' Shiny App interface. On the left, there is a section for 'Upload raw sequencing file' with a text box indicating accepted formats: '.fastq', '.fq', '.fastq.gz' and '.fq.gz'. Below this is a 'Browse...' button and a status 'No file selected'. Further down is a 'Parameter' section with a slider for 'Read length cutoff quantile' ranging from 0.2 to 0.9, with a current value of 0.75. On the right, there are two tabs: 'Output Table' and 'Summary Plot'. Below the tabs, the title 'Top15 flanking genomic regions of insertion sites' is displayed, followed by a note: 'With the top flanking genomic region sequences, please use UCSC Genome Browser blast tool to further locate the insertion site in the genome.'

- 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.

This screenshot shows the 'TransTag' Shiny App interface with the 'Output Table' tab selected. The 'Upload raw sequencing file' section now shows a file named 'H2AZmCherry\_Tol2\_R1.subsample\_10M.fastq' uploaded, with an 'Upload complete' button. The 'Read length cutoff quantile' slider remains at 0.75. The 'Top15 flanking genomic regions of insertion sites' section displays a table of the top 15 flanking genomic regions. The table has two columns: 'kmers' and 'count'.

kmers	count
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCCCGCAGAAAA	163
TCGCCAGGGAACCTGAGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCCCGCAGAAAA	3
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCCCGCAGAAAA	2
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCCCGCAGAAAA	2
TCGCCAGGGAAGTCTGATCGTCCAGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCCCGCAGAAAA	1
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCAGAGTGTGTC	1
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCAGAGTGTGTC	1
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCAAATGGTTTACACATCCCTATCTCTT	1
TCGCCAGGGAAGTCTGATCGTCCCGTGGGTGAAGCCTTGCTGGAGAATCCCAATGGTTTACACATCCCGCAGAAAA	1
TCTGAACCTCTGCAAACTCTGACGTGTGTGGAAGAGAGTGGATATGGGTGTCTGAAAAATATTAACTATTGTA	1

- 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.

**TransTag**

Upload raw sequencing file  
Accepted file formats are '.fastq', '.fq', '.fastq.gz' and '.fq.gz'

Browse... H2AZmCherry\_Tol2\_R1.subsample\_10M.fastq  
Upload complete

Parameter  
Read length cutoff quantile  
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9  
Read size cutoff: 67

Output Table Summary Plot

**Top15 flanking genomic regions of insertion sites**  
With the top flanking genomic region sequences, please use UCSC Genome Browser blast tool to further locate the insertion site in the genome.

kmers	count
TCGCCAGGGAACCTGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	182
TGAGTAAATTTTCCTAAGTACTTGACTTTCACCTTGAGTAAATTTTGAGTACTTTTACACCT	10
TCGCCAGGGAACGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	5
TCGCCAGGGAACCTGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	4
TCGCCAGGGAACCTGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	1
TCGCCAGGGAACCTGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	1
TCGCCAGGGAACCTGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	1
TCGCCAGGGAACCTGAGATCGTCCCGGTGGTGAAGCCTTGCTGGAGAATCCCAAAATGGTTTACAC	1

- 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>).

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

**Zebrafish (danRer11) BLAT Results**

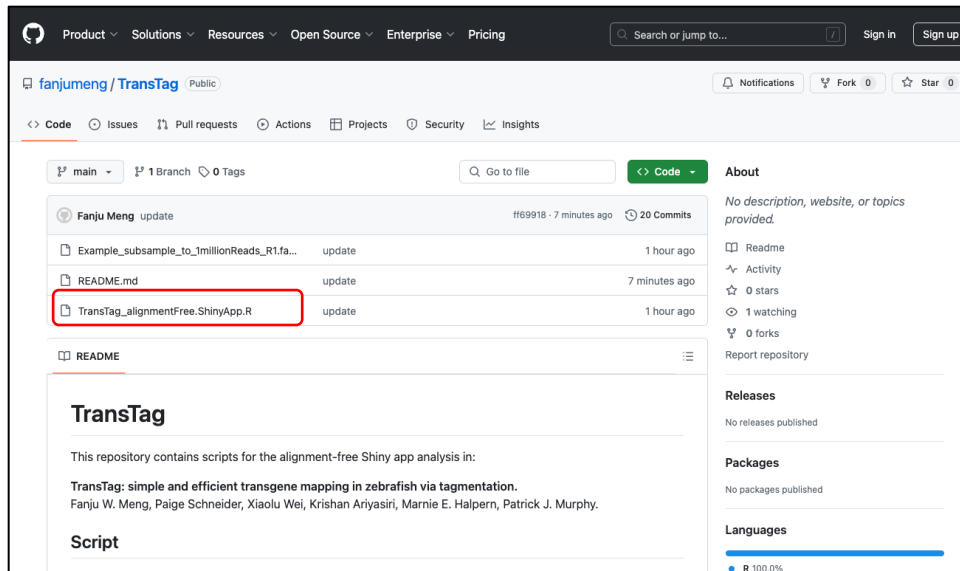
**BLAT Search Results**  
Go back to [chr6.43426669-43433274](https://genome.ucsc.edu/chr6.43426669-43433274) on the Genome Browser.

Custom track name: blat YourSeq  
Custom track description: blat on YourSeq  
Create a stable custom track with these results

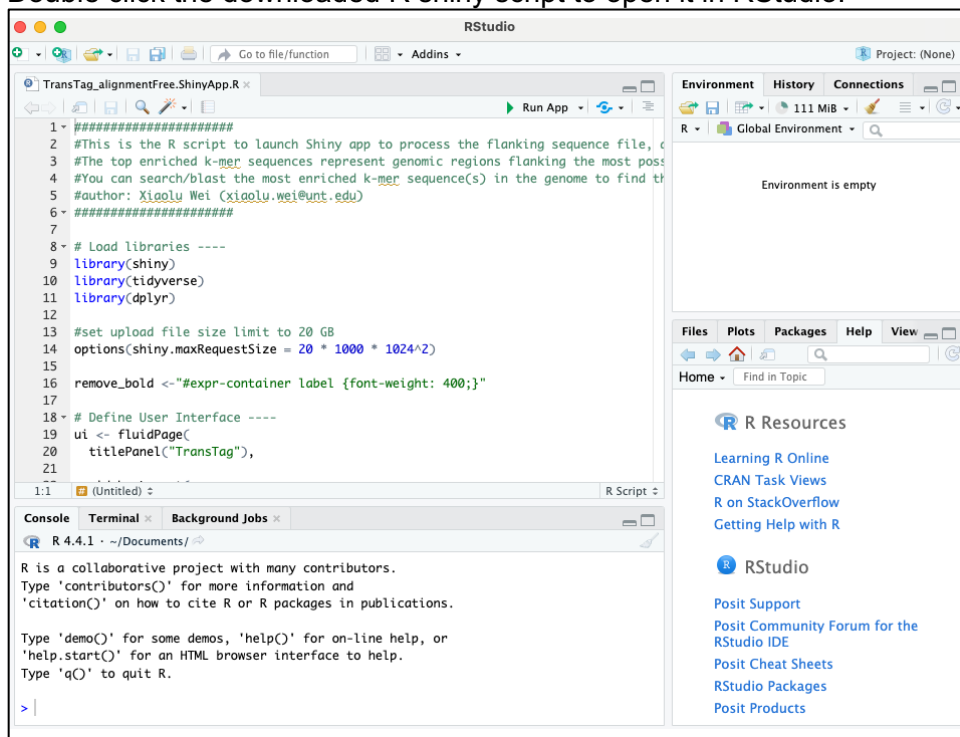
ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser</a> <a href="#">new tab</a> <a href="#">details</a>	YourSeq	80	1	80	80	100.0%	chr23	-	41754070	41754149	80
<a href="#">browser</a> <a href="#">new tab</a> <a href="#">details</a>	YourSeq	21	31	51	80	100.0%	chr23_KZ115697v1_alt	-	75474	75494	21

[What is chrom alt?](#)

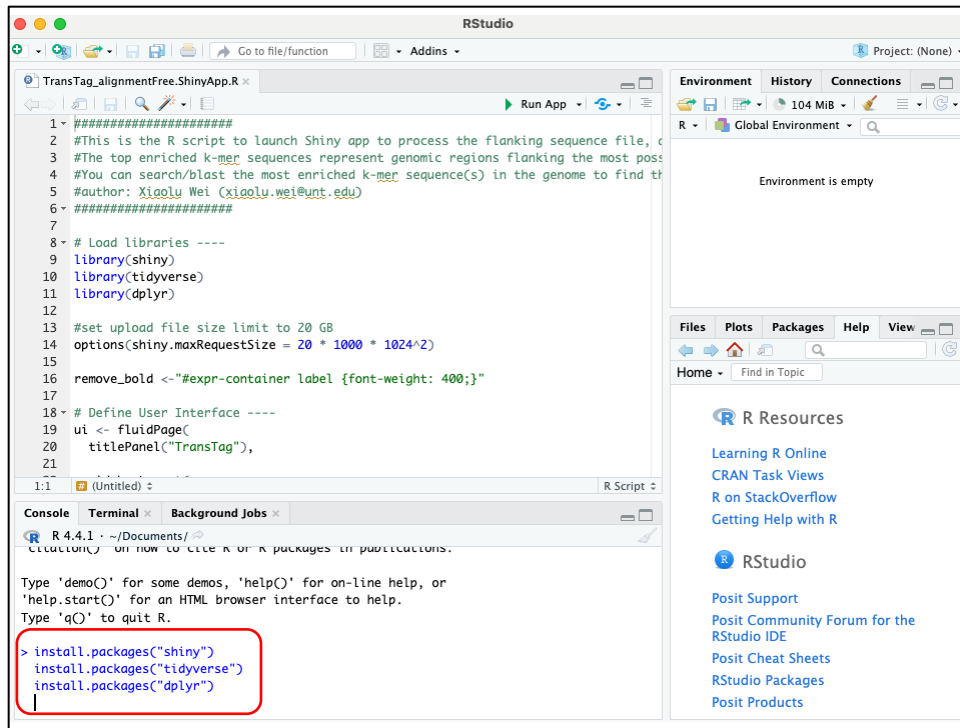
- **Running the shiny script on local computers** (for any size input files)
  - Download and install R and RStudio on local computer following instruction: <https://posit.co/download/rstudio-desktop/>
  - Download alignment-free R Shiny app code "TransTag\_alignmentFree.ShinyApp.R" from GitHub at <https://github.com/fanjumeng/TransTag>.



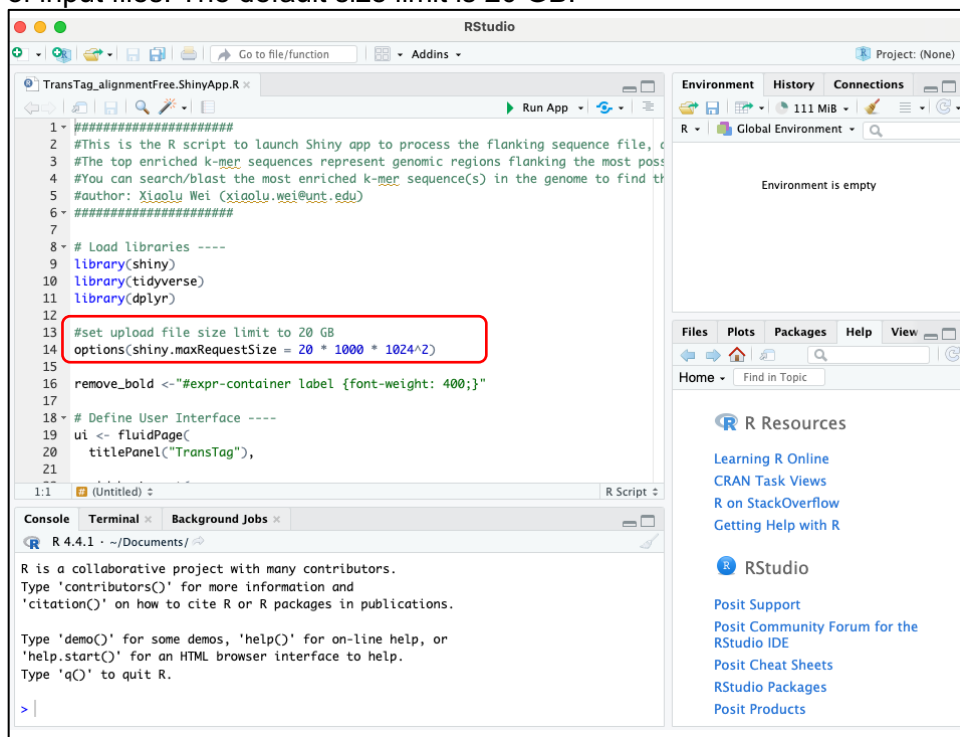
- c. Double click the downloaded R shiny script to open it in RStudio.



- 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.

