Welcome to the ASOs Design Tool user guide. This tool is powered by R Studio Server and runs on a Linux server, providing a reliable and efficient environment for your design tasks. This guide will introduce you to the tool's features and how to use them effectively in our server setup.

This tool is for identifying target regions for antisense oligonucleotides. The script processes genetic sequences, predicts RNA accessibility, determines binding dynamics, and evaluates variants within the human genome. It applies filtering and clustering analyses to efficiently select potential target areas for therapeutic or research purposes.

Table of contents

[Working environment](#_heading=h.30j0zll)

[How does the code work](#_heading=h.i65nvdpibwnv)

[The script in dept](#_heading=h.cgl9sjhbwgmv)

[Function Descriptions](#_heading=h.cwykxlsoqq2d)

[Dependencies](#_heading=h.qtta80ue9db0)

[Parameters and Variables](#_heading=h.l0f9no41z2jk)

[Data Flow](#_heading=h.63yltdo6y8zk)

[Version Control](#_heading=h.vw3vaak83d5a)

[Access to Databases and Tools](#_heading=h.j7qmerb3z226)

[Ensembl Database](#_heading=h.gku9qpeq1tcp)

[TxDb (Transcript Database)](#_heading=h.6ad2nzk3r66c)

[External Tools](#_heading=h.16dr962ywct3)

[ViennaRNA Package](#_heading=h.bapmumvlzzya)

[How do I install the software and its dependencies](#_heading=h.2et92p0)

[Common problems](#_heading=h.tyjcwt)

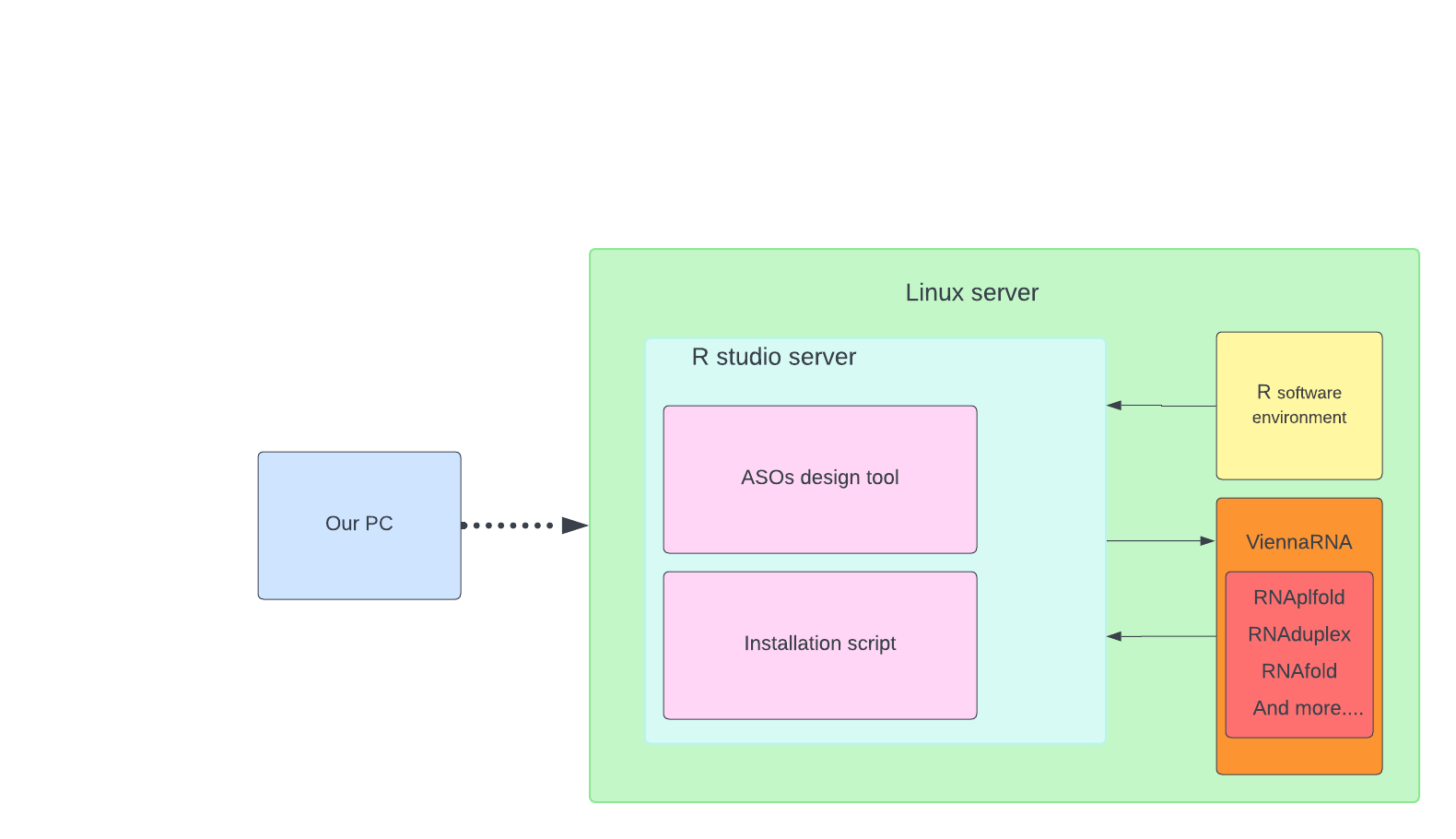
[How to update the databases](#_heading=h.1t3h5sf)

[Updating the Ensembl version](#_heading=h.2s8eyo1)

[FAQ](#_heading=h.17dp8vu)

# 

# Working environment

A small diagram of how this works is shown as follows:  


In our configuration, we utilize a Linux server that is connected to our windows PC via SSH.

SSH, which stands for Secure Shell, is a network protocol that provides administrators with a secure way to access a remote computer.

This server hosts the R Studio Server, which contains two key components: the "ASOs Design Tool", a specially developed R script for creating ASOs, and the "Installation Script", which includes multiple scripts for updates and necessary installations.

The operation of the R Studio Server depends on the "R Software Environment".

Additionally, we have integrated ViennaRNA, an external tool used for predicting RNA structures. This tool is directly accessible from the R Studio Server and provides various functionalities for our research purposes.

# How does the code work

Setup and Library Imports

Initially, the script imports necessary R libraries that provide functions for handling genomic data, performing statistical analysis, and conducting data manipulation. These libraries include tools for working with genomic coordinates (GenomicFeatures, Biostrings), accessing annotation databases (AnnotationDbi, biomaRt), and manipulating data frames in a tidy format (tidyverse, specifically dplyr).

Input Specification

The user specifies the path to a Transcript DataBase (TxDb) and the Ensembl ID of the gene of interest. The Ensembl version is also defined to ensure that the script fetches data from a consistent version of the Ensembl database.

Defining Target Regions

The script then loads a TxDb object, filters gene information to include only those genes present on specified chromosomes, and retrieves the sequence of the gene of interest. It also prepares the data for polymorphism analysis by organizing chromosomal coordinates.

Ortholog and Polymorphism Retrieval

Using biomaRt, the script identifies the mouse ortholog for the human target gene to study conservation across species. It also retrieves information about human polymorphisms associated with the gene, focusing on high-frequency polymorphisms.

Accessibility and Toxicity Calculations

The script employs ViennaRNA's RNAplfold to estimate transcript accessibility at single-nucleotide resolution and RNAfold and RNAduplex to predict secondary structures and duplex formation energies of oligonucleotides, respectively. A custom function calculates the acute neurotoxicity score based on the composition of nucleotides in each potential target region.

Selection Criteria Application

A series of filters are applied to the potential target regions to narrow down the candidates based on gene accessibility, the presence of high-frequency polymorphisms, conservation across species (not conserved in mouse), and the predicted toxicity score.

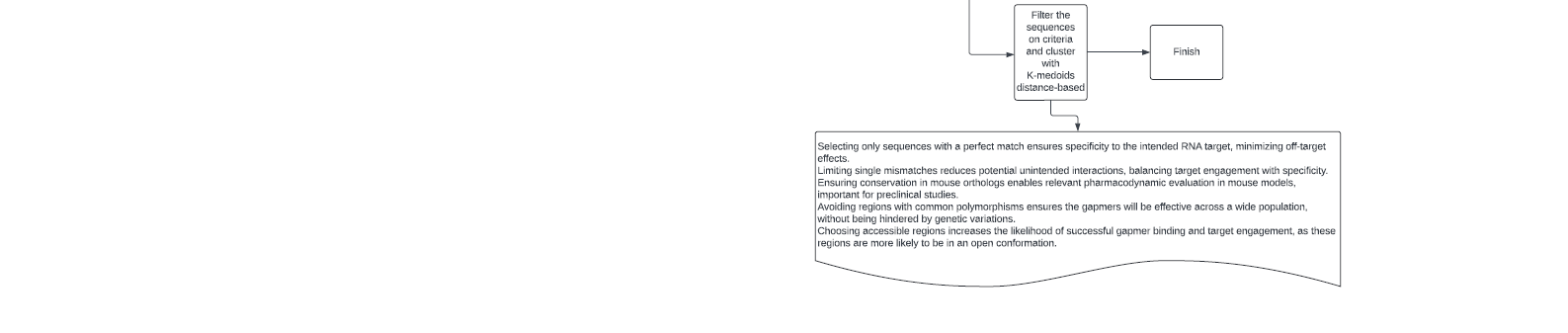
Clustering and Final Selection

The script uses the clara function from the cluster library to perform clustering on the potential target regions based on their start positions. This step aims to group closely located targets and select a representative from each cluster to ensure diverse target coverage along the gene.

Output Generation

Finally, the script writes two CSV files: one containing all filtered target regions and another with the clustered and selected targets. Each file is named with a timestamp to differentiate between different runs or genes analyzed.

# 



# The script in dept

This script is designed for selecting potential target regions within a specific gene's pre-mRNA sequence for the purpose of designing gapmers. Gapmers are oligonucleotides used in gene knockdown studies to inhibit gene expression. The script identifies target regions based on various criteria, including gene accessibility, sequence conservation, polymorphism frequency, and predicted oligonucleotide toxicity.

## Function Descriptions

* filter\_target\_regions(df): Filters target regions based on predefined conditions (e.g., gene hits, accessibility, PM frequency, toxicity score). Returns a dataframe of filtered regions.
* calculate\_acute\_neurotox(xx): Calculates an acute neurotoxicity score for given nucleotide sequences. Returns numeric toxicity scores.
* RNAplfold\_R(seq.char, L.in, W.in, u.in): Estimates RNA target accessibility using the RNAplfold tool. Returns a dataframe of accessibility scores.
* RNAduplex\_R(seqs): Predicts duplex formation energies of oligonucleotides using RNAduplex. Returns numeric energy values.
* RNAselffold\_R(seqs): Predicts self-folding energies of oligonucleotides using RNAfold. Returns numeric energy values.

## Dependencies

This script requires the following R packages and external tools:

* R Packages: BSgenome.Hsapiens.NCBI.GRCh38, GenomicFeatures, AnnotationDbi, BiocManager, cluster, tidyverse, biomaRt, Biostrings, dplyr
* External Tools: ViennaRNA package tools (RNAplfold, RNAduplex, RNAfold)

## Parameters and Variables

* path: Path to the database/TxDb file.
* ensembl\_ID: The Ensembl ID of the gene of interest.
* ensemblversion: Version of the Ensembl database to use.
* oligo\_lengths: Range of oligonucleotide lengths to consider for target regions.

## 

## 

## Data Flow

1. Setup and Initialization: Load necessary libraries and set parameters.
2. Database Loading and Filtering: Load the TxDb object and filter for the gene of interest based on the specified Ensembl ID.
3. Ortholog and Polymorphism Retrieval: Use biomaRt to fetch orthologous genes and polymorphisms.
4. Sequence Analysis: Estimate transcript accessibility, predict duplex and self-folding energies, and calculate toxicity scores.
5. Target Selection: Apply filters to select potential target regions based on defined criteria.
6. Output Generation: Write the results to CSV files for further analysis or experimental validation.

## Version Control

A backup of the script are found on the github “<https://github.com/MajatoProj/ASOstool>”

# Access to Databases and Tools

### Ensembl Database

* Description: Ensembl provides a comprehensive and integrated source of annotation of, mainly, vertebrate genome sequences.
* Access: The script uses the biomaRt R package to query Ensembl data. Ensure that you have specified the correct version of the Ensembl database using the ensemblversion variable in the script.
* Installation: biomaRt can be installed via Bioconductor with the command BiocManager::install("biomaRt").

### TxDb (Transcript Database)

* Description: TxDb objects provide a unified structure for storing annotation data. They can be sourced from various locations, including Bioconductor.
* Access: For this script, the path to a TxDb object specific for human genes (HSapiens) is required. This path is specified in the path variable.
* Installation: TxDb objects for various organisms and genome builds can be installed from Bioconductor. For instance, TxDb.Hsapiens.UCSC.hg38.knownGene can be installed using BiocManager::install("TxDb.Hsapiens.UCSC.hg38.knownGene").
* Usage: Load the TxDb object using loadDb function from the GenomicFeatures package.

## 

## External Tools

### ViennaRNA Package

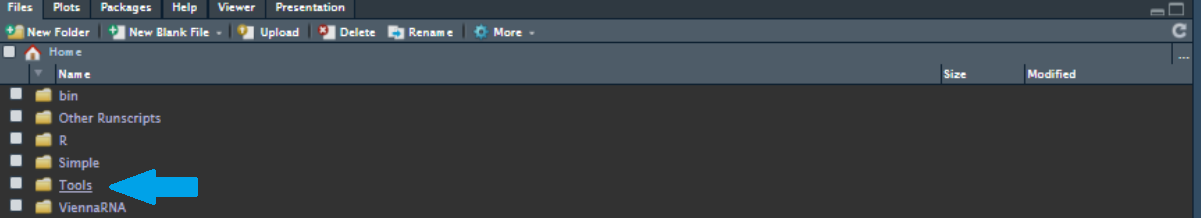
* Description: The ViennaRNA Package consists of a C library and several command-line tools for the prediction and comparison of RNA secondary structures.
* Tools Used: RNAplfold, RNAduplex, RNAfold.
* Installation:
  + Download the latest version from the ViennaRNA website.
  + Follow the installation instructions provided, which typically involve configuring and compiling the source code.
* Usage: The script calls these tools via system commands in R. Ensure they are installed in a location that's included in your system's PATH variable, or modify the script to include the full path to the tools.

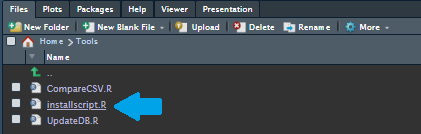
# How do I install the software and its dependencies

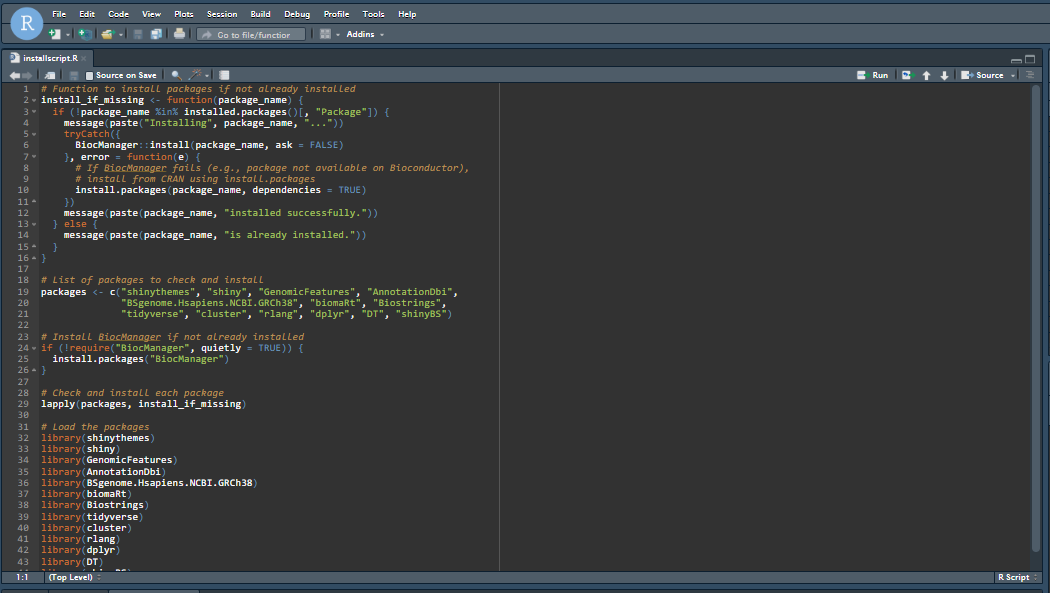
**Updating the packages is not necessary to keep the tool working, so only use this when needed!**

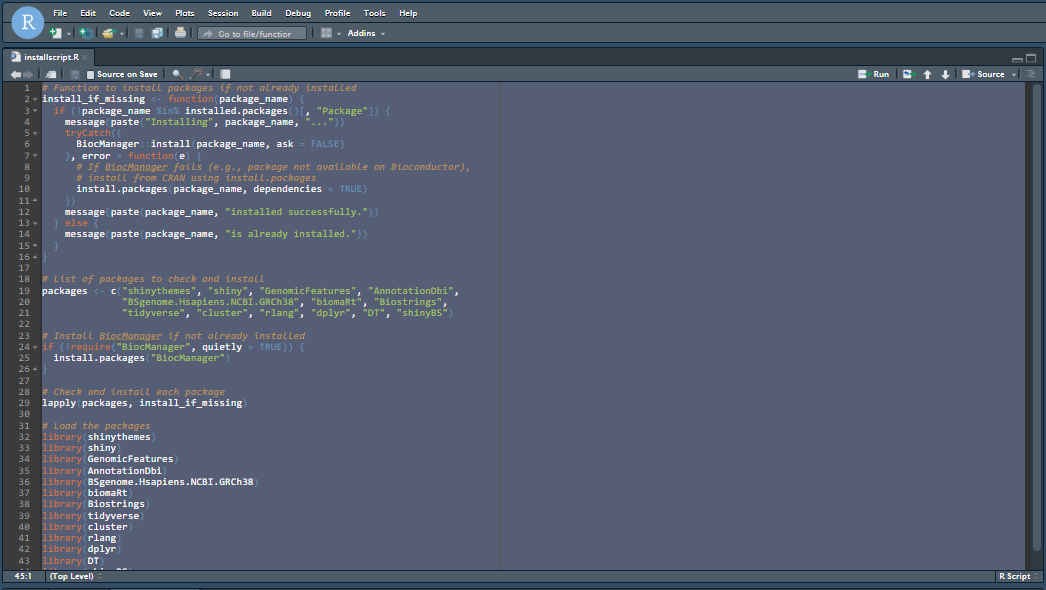
Installing and updating the dependencies can be done with the “installscript.R” inside the Tools folder which can used inside of the R studio environment:

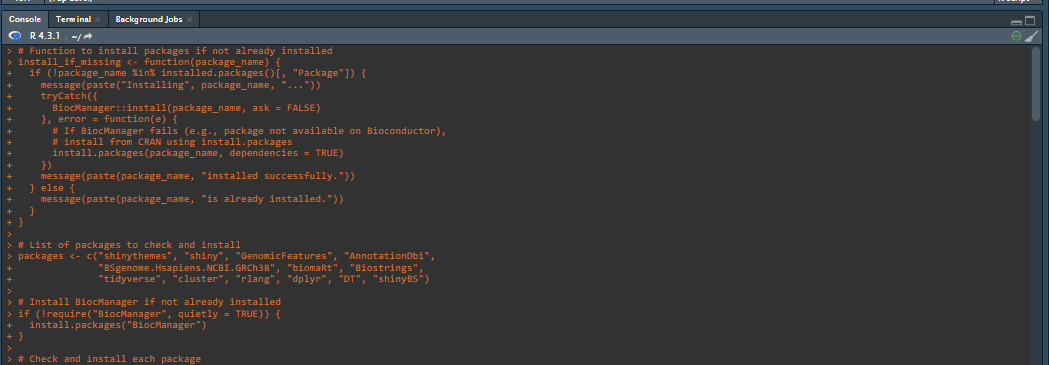
1) Simply open the script in the lower right.



  
2)When the script is opened select all the lines using the button combination: “Ctrl+A” on Windows/linux and “Commmand-A” on Mac.



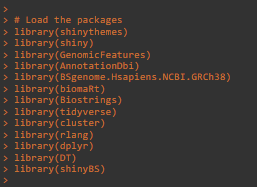
  
3)Then press “Shift+Enter”



The terminal might look different.

4)The script will now install all the dependencies needed,

5) The packages will then be tested by being loaded in



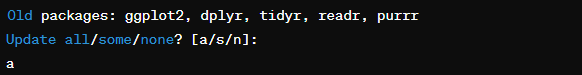
6) It will print “done” when the script is finished



## Common problems

A package update could return an error, most of the time this is due to one of the following problems:

* There needs to be more storage space for the update.
  + Try clearing up some space and try again.
* There is an update prompt asking if you want to update all/some/none of the packages.



* + Try to respond with all if it doesn't work. try to update the packages by hand using the command: update.packages(“name of package”)

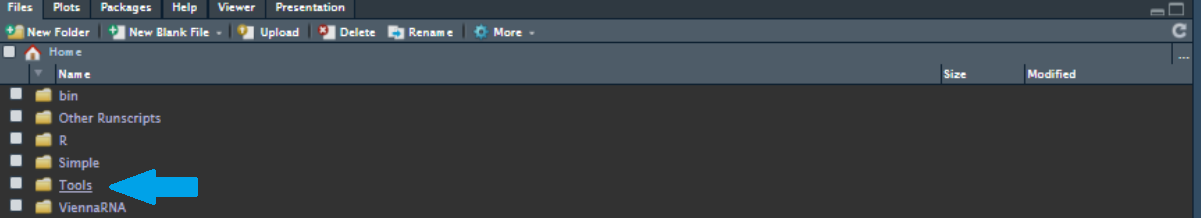
# 

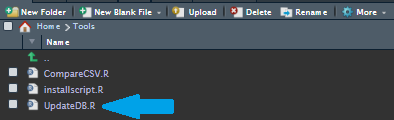
# How to update the databases

Updating the databases can be a little more difficult.

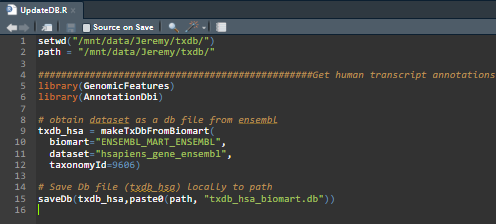
To update the DB file you can use the “UpdateDB.R” script in the Tools folder.  
**Make sure to backup the old db file before removing it from the original map.**simply run the script as you would do as the “installscript.R”

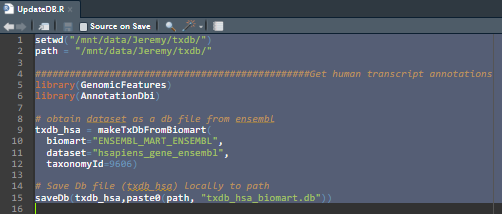
1) Simply open the script found in the Tools folder in the lower right.



  
2)When the script is opened select all the lines using the button combination: “Ctrl+A” on

Windows/linux and “Commmand-A” on Mac.





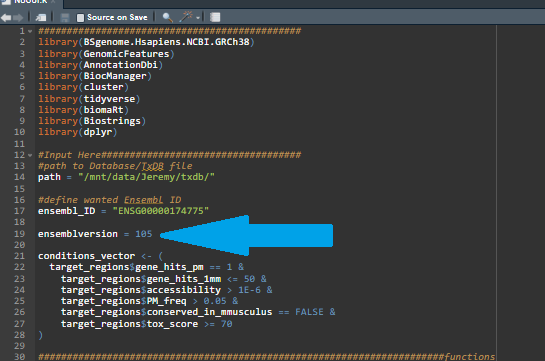
3)Then press “Shift+Enter”, the script is finished when “done” is printed in the console.



4)Make sure everything still works.

## Updating the Ensembl version

**is not advised without doing deep research, some versions are missing crucial information** for the tool. A stable version is 105.  
To change this, go into the script and search for the following line:

**

Here you can change the version to whatever is preferred.

As of April 5th, 2024, only version 105 has been tested for a stable version.

Other versions may result in an incorrect working script.

# 

# FAQ

**Q1: What do I need to start using the system?**

A1: You need access to a computer with an Erasmus MC internet connection to connect to the Linux server.

**Q2: How do I connect to the Linux server?**

A2: Use your SSH client with the server's IP address and your credentials. Detailed steps can be found in the "How to use the tool" section of the user guide.

**Q3: What is the ASOs Design Tool?**

A3: The ASOs Design Tool is a custom R script available within the R Studio Server environment. It helps find areas in genes where antisense oligonucleotides (ASOs) can be used, offering suggestions for possible ASOs to assist in research.

**Q4: How do I update the system or install new tools?**

A4: Updates and installations are managed through the "Installation Script." Refer to the "How to update the databases" & “How do I install the software and its dependencies” sections for step-by-step instructions.

**Q5: Do I need any specific software to use R Studio Server?**

A5: Yes, the R Studio Server requires the R Software Environment to function. This environment is pre-installed on the Linux server.

**Q6: What is ViennaRNA and how do I use it?**

A6: ViennaRNA is an external tool integrated into our system for RNA structure prediction. It's accessible from the R Studio Server for conducting various analyses.

Troubleshooting ViennaRNA is best done by contacting the developer's email:

“rna@tbi.univie.ac.at”

**Q7: Can I access the system from any PC?**

A7: Yes, as long as you’re connected to the Erasmus MC internet, have SSH access, and the necessary credentials, you can connect to the Linux server from any PC.