# **Retrieving TargetScanMouse data**

We downloaded UTR sequence from TargetScanMouse using the link below:

<http://www.targetscan.org/cgi-bin/targetscan/data_download.mmu80.cgi>

**Data Download / Gene and miRNA annotations**

**File:** **UTR Sequences - (377.8 MB)**

**Description:** UTR sequences from 46-way alignments

**Fields:** Transcript ID, Gene ID, Gene Symbol, Species ID, and UTR Sequence

**# of rows: 1,294,698**

**We filter Species ID column for *Mus musculus* (mouse) (taxonomy id: 10090). After filtering, we obtain 24898 rows.**

# **Input Genes**

We use 22301 ENSMUSG genes. (macrophage\_riboseq\_hkazan\_genes).

# **Motif List**

We use the following motifs:

MOTIF\_LIST = [  
 "AUUUA",  
 "UAUUUAU",  
]

# **Counting the motifs in sequences**

For each UTR sequence, we count the non-overlapping occurrences of each motif.

# **Analysis results**

For each input gene, we count the occurrences of motifs in the UTR sequences.

Among 22301 genes, UTR sequence is not found for 2031 genes.

In case of finding several UTR sequences for the same ENSMUSG, we take the longest one (after all of them converted to non-dash form).

Further, we normalize the occurrence count of each motif by the length of the UTR sequences.