3.4.1 - Getting Started

- Searching protein domains on alternatively spliced regions of human gene TNNT1
  - Input
  - Command-lines

Searching protein domains on alternatively spliced regions of human gene TNNT1

According to RefSeq (NM_003283),

> *This gene encodes a protein that is a subunit of troponin, which is a regulatory complex located on the thin filament of the sarcomere. This complex regulates striated muscle contraction in response to fluctuations in intracellular calcium concentration.*

Input

- Annotation of eight alternative transcripts from GENCODE Basic v24 (Download)
- Chromosome 19 FASTA file from **GRCh38/hg38 (**Download)
- Reference file (Download)
- HMM file (Download)

Command-lines

Obtaining reference transcript sequence

```
$> astalavista -t astafunk --tref --gtf tnnt1.gtf --genome ~/example/genome/ > reference_tx.fasta
```

Creating reference file

```
$> hmmsearch --domtblout reference_file ~/Databases/Pfam/Pfam-A.hmm reference_tx.fasta
```

Obtaing a reduced HMM file

```
$> grep -v "#" reference_file | awk '{print $5}' | sort | uniq | hmmfetch -f ~/Pfam/Pfam-A.hmm - > database.hmm
```

or skip these commands and use directly the whole database Pfam-A.hmm as parameter for the option [–hmm].

Running AstaFunk to obtain alternatively spliced domains

```
astalavista -t astafunk --genome ~/example/genome/ --gtf tnnt1.gtf --reference reference_file --hmm database.hmm
```

Description of the output columns can be found in 3.4 - Tool ASTAFUNK.