TAGOOS Documentation

Release 0.2.1

Aitor GONZALEZ

Aug 26, 2022

Contents:

1 UCSC genome browser tracks **1**

2 Scores and annotations **3**
- 2.1 Remote access . . . 3
- 2.2 Local access . . . 3

3 Download **5**

4 What's New in 0.2.0 (April XX, 2019) **7**

5 What's New in 0.2.0 (April XX, 2019) **9**

6 Indices and tables **11**

CHAPTER 1

UCSC genome browser tracks

We provide UCSC tracks here:

- TAGOOS scores in the hg19 UCSC browser
- TAGOOS scores in the hg38 UCSC browser

When you arrive to the UCSC genome browser, there are four TAGOOS related tracks:

- **TAGOOSintronNegLog10pval** : The negative logarithm with base 10 of the P-value for the intronic TAGOOS scores
- **TAGOOSintergNegLog10pval**: The negative logarithm with base 10 of the P-value for the intergenic TAGOOS scores

CHAPTER 2

Scores and annotations

2.1 Remote access

Download and install TABIX. Then retrieve scores and annotations remotely in the command line. For instance for the SNP rs227727 (hg19, chr17:54776955-54776955)

```
tabix http://pedagogix-tagc.univ-mrs.fr/~gonzalez/tagoos/release/180328/tagoos_
↪intergenic_hg19.bed.gz chr17:54776955-54776955
```

Output columns are described here

2.2 Local access

If you plan to retrieve score often, it is better to download BED.GZ and TBI files here and use TABIX locally. The same example with a local file:

```
tabix tagoos_intergenic_hg19.bed.gz chr17:54776955-54776955
```

CHAPTER 3

Download

Gzipped BED (bed.gz) and TABIX (bed.gz.tbi) files can be downloaded here: TAGOOS files

Table 1: TAGOOS BED and TBI files

Size	Last modified	File name
1.2G	May -26-2018	tagoos_intergenic_hg19.bed.gz
685K	May -26-2018	tagoos_intergenic_hg19.bed.gz.tbi
1.2G	May -26-2018	tagoos_intergenic_hg38.bed.gz
684K	May -26-2018	tagoos_intergenic_hg38.bed.gz.tbi
2.0G	May -26-2018	tagoos_intronic_hg19.bed.gz
754K	May -26-2018	tagoos_intronic_hg19.bed.gz.tbi
2.0G	May -26-2018	tagoos_intronic_hg38.bed.gz
754K	May -26-2018	tagoos_intronic_hg38.bed.gz.tbi

Column descriptions of bed files

1. Chromosome number with "chr" suffix. It goes from 1 to 22, i.e. no X o Y chromosomes
2. Starting position (zero-based)
3. End position (one-based)
4. Score
5. p-value
6. Negative log with base 10 of the p-value
7. Comma-separated list of annotation variables

Example line for the rs227727 SNP (chr17:54776954-54776955 in hg19 genome) retrieved with tabix

```
chr17	54776954	54776955	0.26526207	0.00279416049738879	2.
↪55374865150445	ahr.mcf7.GSE41820ReMap,atf3.NS.ENCODEReMap,atf7.mcf7.encode2,
↪dnase.bodyofpancreasmaleadult_54years_.encode2,dnase.brainmalefetal_72days_
↪andmalefetal_76days_.encode2,dnase.bronchialepithelialcell.encode2,dnase.daoy.
↪encode2,dnase.esdr_h1_bmp4_meso.roadmap,dnase.fibroblastofupperlegskinmalefetal_
↪12weeks_.encode2,dnase.gastrocnemiusmedialismaleadult_54years_.encode2,dnase.
↪heartfetal_96days_.encode2,dnase.ipsdf19.11malenewborn.encode2,dnase.
↪largeintestinemalefetal_108days_.encode2,dnase.leftkidneyfemalefetal_87days_.
↪encode2,dnase.mg63.encode2,dnase.mus_trnk_fet.roadmap,dnase.skinfibroblastmalefetal_
↪97days_.encode2,dnase.smallintestinemalefetal_108days_.encode2,dnase.
↪stomachmalechild_3years_.encode2,dnase.testismalefetal.encode2,dnase.
↪thyroidglandmaleadult_37years_.encode2,dnase.trophoblastcellfetal_21week_.encode2,
```

(continues on next page)

(continued from previous page)

CHAPTER 4

What's New in 0.2.0 (April XX, 2019)

- Fixed ReadTheDocs

CHAPTER 5

What's New in 0.2.0 (April XX, 2019)

- Manuscript was accepted in Nucleic Acid Research
- Merged source code and ReadTheDocs documentation

TAGOOS is a nucleotide scoring tool for non-coding (Intronic and intergenic) regions. There are two underlying models trained with the XGBOOST algorithm using intronic and intergenic associated SNPs (GWAS P-value $< 5 \cdot 10^{-8}$) from the GRASP database. The predictive variables have been selected by the learning algorithm among 4684 gene regulation related annotations such as histone modifications, eQTLs or transcription factors in different tissues from these databases:

- Expressed enhancers
- ENCODE
- GTEx
- ReMap
- RoadMap
- H3K27ac from the Young lab

The scores can be downloaded or accessed remotely using:

- UCSC (Remote)
- Tabix (Remote and local)

CHAPTER 6

Indices and tables

- genindex
- modindex
- search