

ATHENA Tutorial

Installation:

- Download the ATHENA source file from <http://ritchielab.psu.edu/software>
- Unzip the tar ball `athena-1.0.tar.gz`
`tar -xvzf athena-1.0.tar.gz`
- `cd athena/src`
- `make`

Note: By Default, ATHENA is not installed with PARALLEL processing. You can turn it ON by editing `src/athena/Makefile`

“Comment out `DFLAGS = -DPARALLEL` and put `#` before `DFLAGS =` in next line”

Files included in the tutorial:

DATASETS:

`/sample_sets/`

DESCRIPTION: Both datasets were generated using genomeSIMLA.

-25 SNPs

-Heritability = 0.1

-Minor allele frequency of 2 functional SNPs = 0.4

-Two functional SNPs are G1 and G10

-1000 cases and 1000 controls

-100 datasets for each model

CONFIGURATION FILE:

makeathenaconfigs.pl

Script will make 100 configuration files that correspond to the 100 datasets for each model. While running the script, please specify the following four options:

- p Prefix for datasets
- s Suffix for datasets
- d Directory containing datasets
- a Algorithm to use (GENN or GESR, must match grammar!)
- g Full path to grammar file

Specifying the output directory is optional:

- o Output directory (optional -- set to dataset directory when omitted)

Command running the sample sets (you can also type `./makeathenaconfig.pl` for USAGE):

```
./makeathenaconfigs.pl -p Sample_Dataset- -s .txt -g  
/path/to/grammar/file -d /path/to/datasets
```

GRAMMAR FILES:

Grammar files describe the grammars for the algorithm. One might choose one of them for the models.

`genn_add.gram` - Grammar File for GENN (Add-only nodes)

`symbolic_regress.gram` - Grammar File for GESR (+, -, /, * functions).

Running ATHENA

If, you want to run ATHENA for single dataset just execute the command:

```
./bin/athena #.config  
(Config file generated from makeathenaconfig.pl)
```

To run ATHENA in parallel (`athenap`), you must use PBS queuing system. A sample PBS script called `SampleAthena.pbs` is included in this tutorial. The script **`run_athenap.pl`** will be used to run `athenap` on all 100 configuration files.

In the **pbs script** `SampleAthena.pbs` specifies the number of processors you want and make sure it matches the number after `-np` in the `run_athenap.pl` script so as to run it more efficiently.

Examples of the two paths in pbs:
`/path/to/run_athenap.pl`
`/path/to/configs/*.config`

Commands for Running ATHENA:

Create Configuration Files

Execute `makeathenaconfig.pl`:

```
./makeathenaconfigs.pl -p Sample_Dataset- -s .txt -d  
/path/to/sample/dataset -a (GENN/GESR) -g  
/path/to/grammar/file/genn_add.gram -o  
/path/to/output/directory
```

Update run_athena.pl and run_athenap.pl

You need to update these to PERL scripts by providing the path to ATHENA executables.

Update PBS Script

Update the PBS script by specify the paths of **run_athenap.pl** and **(*config)**configuration files directory

```
vi SampleAthena.pbs
```

Run ATHENA

Execute PBS script **SampleAthena.pbs**:

```
qsub SampleAthena.pbs
```

RESULTS:

By default configuration, ATHENA generates following result files:

- *.athena.sum (Summary File)
- *.best (Best model for each cross validation)
- *.dot (Dot compatible file for visualization)

athena_power.pl summarizes results from *.sum files after ATHENA is run. Execute **athena_power.pl** to see the usage:

```
./athena_power.pl
```

This script summarizes the number of times either or both of the functional SNPs for example G1,G10 were present in your best model. Run in the directory where your summary files (*.athena.sum) are located.

Try on your own:

It is recommend playing with parameters in configuration file like popSize and number of generations to compare the power of ATHENA to find either one or both of the functional SNPs in the best model.

It should be noted that due to the nature of GENN programming, running multiple times may yield different results.