

# Transcriptome-Wide Association Study Pipeline (TWAS-pipeline)

## INSTALLATIONS

### --- TWAS ---

The following packages by the developers

Package	URLs
TWAS	<a href="http://sashagusev.github.io/TWAS/">http://sashagusev.github.io/TWAS/</a>
weight files	<a href="https://data.broadinstitute.org/alkesgroup/TWAS/">https://data.broadinstitute.org/alkesgroup/TWAS/</a>
z-score clean program	<a href="https://data.broadinstitute.org/alkesgroup/TWAS/ETC/CLEAN_ZSCORES.tar.bz2">https://data.broadinstitute.org/alkesgroup/TWAS/ETC/CLEAN_ZSCORES.tar.bz2</a>

are required and be unpacked. In addition, lists of genes in the three populations are made through the following scripts,

```
TWAS=/genetics/bin/TWAS
cd $TWAS
for pop in MET NTR YFS
do
  ls WEIGHTS_$1 | sed 's\\/\\g' > $pop.lst
done
```

--- **TWAS-pipeline** ---. The pipeline is installed as follows,

```
git clone https://github.com/jinghuazhao/TWAS-pipeline
```

On our system, `TWAS.sh` and `TWAS_get_weights.sh` for TWAS and `twas.sh`, `twas2.sh`, `twas2-collect.sh` and `twas2-1.sh` for TWAS-pipeline have symbolic links under `/genetics/bin` and available from the `$PATH` environment. A [Stata](#) equivalent has been developed by Dr Jian'an Luan.

To accommodate the suggestion of p value in accordance with the Z-score in the output, `pnorm.c` is included which can be compiled as follows,

```
gcc pnorm.c -lm -o pnorm
```

and a call `pnorm z_score` yields a p value with more decimal places.

--- **GNU Parallel** ---. Further information is available from [here](#).

## RUNNING THE PIPELINE

Suppose you have a file containing GWAS summary statistics, you can run the pipeline as follows,

```
twas.sh input_file
```

where the `input_file` is in tab-delimited format containing SNP\_name, SNP\_pos, Ref\_allele, Alt\_allele, Beta and SE. The output will be contained in `<input file>.imp`.

This assumes that ssh can access nodes in a clusters freely and in case this has not been done, a single node mode is more appropriate,

```
twas-single.sh input_file
```

## TWAS USING GTEx

This is achieved with `gtex.sh` and `gtex.subs` using weights from the GTEx project. File `GTEx_WEIGHTS.bim.gz` was created by `GTEx_WEIGHTS.sh` in accordance with reference `.bim` files used by TWAS and `GTEx_WEIGHTS.lst` was created to facilitate the imputation.

## ADDITIONAL TOPICS

These are available from [the wiki directory](#),

- [BUILDING REFERENCE PANEL](#)
- [AN EXPOSITION WITH GIANT DATA](#)
- [EPIGENOMEWIDE ASSOCIATION](#)
- [FUSION pipeline](#)

## ACKNOWLEDGEMENTS

The work is possible with an EWAS project within the MRC Epidemiology Unit, for which colleagues and collaborators have contributed.

## REFERENCES

Gusev A, et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, 48, 245-252