
HUNT-GWAS-Pipeline Documentation

Endre Bakken Stovner

Jun 15, 2018

1	Installation	3
1.1	Singularity	3
2	Quick start	5
3	Indices and tables	7

The HUNT-GWAS-Pipeline is a highly configurable pipeline that runs GWAS analyses, plots the results and creates ample logs to help you understand which steps were performed. No programming skill is required; the only thing the pipeline needs is a filled out configuration file.

The HUNT-GWAS pipeline leverages the innovative Snakemake workflow language to ensure reproducibility and scalability at no extra cost or additional complexity for the user.

To ensure the reproducibility across machines and clusters, each step in the workflow uses an accompanying Singularity-container with the necessary software installed.

The workflow has ample tests to quickly allow for continual innovation and contributions without risking introducing new bugs.

CHAPTER 1

Installation

The HUNT-GWAS Pipeline requires Snakemake and the conda package manager. Singularity is optional, but recommended. This page will show you how to install them.

First we will install the conda package manager. Go to <https://www.anaconda.com/download/> and select the Python 3+ version.

Then follow the installation instructions.

Next, we will need to install snakemake. We will do this using anaconda.

```
# install snakemake from the channel bioconda  
conda install -c bioconda snakemake
```

Finally you can install the HUNT-GWAS Pipeline with the command

```
git clone git@gitlab.com:huntgenes/gwas_pipeline.git
```

1.1 Singularity

Singularity is a way to package software and the operating system it is run on to ensure complete reproducibility. Using it with the HUNT-GWAS pipeline is highly recommended. The installation-instructions are here: <http://singularity.lbl.gov/install-linux>

CHAPTER 2

Quick start

If you have downloaded the HUNT-GWAS pipeline, go to the folder where it is located.

```
# git clone git@gitlab.com:huntgenes/gwas_pipeline.git  
cd gwas_pipeline
```

The GWAS analysis requires a configuration-file and dosage files to run. The configuration file is where the user defines the settings for the pipeline, and the genetic data are the files where we should look for statistically significant SNPs.

In our case, the

CHAPTER 3

Indices and tables

- `genindex`
- `modindex`
- `search`